MANUAL OF ANALOGUE SOUND RESTORATION TECHNIQUES

by Peter Copeland

The British Library

MANUAL OF ANALOGUE SOUND RESTORATION TECHNIQUES

by Peter Copeland

This manual is dedicated to the memory of

Patrick Saul

who founded the British Institute of Recorded Sound,*
and was its director from 1953 to 1978,
thereby setting the scene which made this manual possible.

Published September 2008 by The British Library
96 Euston Road, London NW1 2DB
Copyright 2008, The British Library Board
www.bl.uk

* renamed the British Library Sound Archive in 1983.

CONTENTS

Preface 1
Acknowledgements 2
1 Introduction 3
1.1 The organisation of this manual 3
1.2 The target audience for this manual 4
1.3 The original sound 6
1.4 Operational principles 8
1.5 A quadruple conservation strategy 10
1.6 How to achieve objectivity 11
1.7 The necessity for documentation 12
2 The overall copying strategy 13
2.1 The problem to be solved 13
2.2 General issues 13
2.3 The principle of the "Power-Bandwidth Product" 14
2.4 Restricting the bandwidth 16
2.5 Deciding priorities 17
2.6 Getting the best original power-bandwidth product 17
2.7 Archive, objective, and service copies 19
2.8 "Partially objective" copies 20
2.9 Documentation strategy 20
2.10 Absolute phase 22
2.11 Relative phase 23
2.12 Scale distortion 23
2.13 Conclusion 24
3 Digital conversion of analogue sound 25
3.1 The advantages of digital audio 25
3.2 Technical restrictions of digital audio - the "power" element 26
3.3 Technical limitations of digital audio: the "bandwidth" element 28
3.4 Operational techniques for digital encoding 30
3.5 Difficulties of "cloning" digital recordings 30
3.6 Digital data compression 33
3.7 A severe warning 35
3.8 Digital watermarking and copy protection 37
3.9 The use of general-purpose computers 38
3.10 Processes better handled in the analogue domain 39
3.11 Digital recording media 40
4 Grooves and styli 42
4.1 Introduction 42
4.2 Basic turntable principles 43
4.3 Pickups and other devices 44
4.4 Conventional electrical pickup considerations 45
4.5 Operational procedure for selecting a stylus 47
4.6 U-shaped and V-shaped grooves 48
4.7 The principle of minimising groove hiss 52
4.8 "Soft" replay styli 53
4.9 "Hard" replay styli 55
4.10 Stereo techniques 57
4.11 "Elliptical" and other styli 59
4.12 Other considerations 61
4.13 Playing records backwards 63
4.14 Half-speed copying 65
4.15 Distortion correction 65

4.16 Radius compensation 67
4.17 Electronic click reduction 70
4.18 Electronic hiss reduction 74
4.19 Eliminating rumble 76
4.20 De-thumping 76
4.21 Future developments 77
4.22 Recommendations and conclusion 78
5 Speed setting 81
5.1 Introduction 81
5.2 History of speed control 82
5.3 History of speed-control in visual media 84
5.4 Setting the speed of old commercial sound records 86
5.5 Musical considerations 90
5.6 Strengths and weaknesses of "standard pitch" 91
5.7 Non-"standard" pitches 91
5.8 The use of vocal quality 92
5.9 Variable-speed recordings 93
5.10 Engineering evidence 95
5.11 Timings 97
6 Frequency responses of grooved media 99
6.1 The problem stated 99
6.2 A broad history of equalisation 100
6.3 Why previous writers have gone wrong 100
6.4 Two ways to define "a flat frequency response" 101
6.5 Equalisation ethics and philosophy 102
6.6 Old frequency records as evidence of characteristics 103
6.7 Two common characteristics 104
6.8 Practical limits to equalisation 106
6.9 Practical test discs 107
6.10 International standard microgroove test discs 107
6.11 Coarsegroove (78rpm) test discs 109
6.12 Generalised study of electromagnetic cutters 111
6.13 Characteristics of "simple" cutterheads 111
6.14 High-resistance cutterheads 115
6.15 Western Electric, and similar line-damped recording systems 116
6.16 Western Electric revolutionises sound recording 117
6.17 The Western Electric microphone 117
6.18 The Western Electric amplifier 118
6.19 Documentation of HMV amplifier settings, 1925-1931 118
6.20 The Western Electric cutterhead 119
6.21 How to recognise recordings made with Western Electric 1A and 1B systems 121
6.22 Summary of equalisation techniques for the above 122
6.23 Western Electric developments after 1931 122
6.24 Recognising recordings made on RCA-modified and Western Electric 1C and 1D systems 124
6.25 Summary of equalisation techniques for the above 124
6.26 Other systems using line-damped cutterheads: British Brunswick, Decca, DGG 1925-1939 124
6.27 Other systems using line-damped cutterheads: The Lindström system 125
6.28 Conclusion to line-damped systems 126
6.29 The Blumlein system 127
6.30 The Blumlein microphone 127
6.31 The Blumlein moving-coil cutterhead 128
6.32 Test discs made by Blumlein's system 129
6.33 How to recognise Blumlein cutters on commercial records, 1931-1944 130
6.34 Summary of how to equalise the above 130
6.35 The Gramophone Company system 130
6.36 How to recognise the Gramophone Company system on commercial records 131
6.37 Summary of how to equalise "Gramophone system" recordings 131
6.38 Extended-Range Blumlein recordings (1943-5) and later systems 131
6.39 Summary of how to equalise "Extended-Range" and subsequent systems 132

6.40 Early EMI long-playing and 45 r.p.m. records....132
6.41 Other systems giving a "Blumlein-shaped" curve - amateur and semi-pro machines....133
6.42 British Decca-group recordings 1935-1944....134
6.43 Summary of how to equalise Decca-group recordings 1935-1944....135
6.44 Synchrophone....135
6.45 Summary of how to equalise Synchrophone recordings....135
6.46 Conclusion to "Blumlein-shaped" equalisation....135
6.47 The Marconi system....136
6.48 BBC Disc Record equalisation - Introduction....137
6.49 The subject matter of these sections....138
6.50 Pre-history of BBC disc recording....139
6.51 BBC matrix numbers....139
6.52 BBC "current library" numbers....140
6.53 BBC M.S.S. recordings....141
6.54 BBC American equipment....142
6.55 BBC transportable equipment....142
6.56 BBC coarsegroove 33rpm discs....142
6.57 Later BBC coarsegroove systems....143
6.58 Early BBC microgroove discs....145
6.59 BBC "CCIR characteristics"....145
6.60 RIAA and subsequently....146
6.61 Brief summary of BBC characteristics....146
6.62 "Standard" equalisation curves....147
6.63 General history of "standards"....147
6.64 Defining standard curves....149
6.65 Discographical problems....150
6.66 General history of changeover procedures....151
6.67 NAB (later "NARTB") characteristics....152
6.68 Decca Group (UK) "ffrr" characteristics....152
6.69 American Columbia LPs....155
6.70 "AES" characteristics....156
6.71 Various RCA characteristics....157
6.72 "CCIR" characteristics....158
6.73 "DIN" characteristics....158
6.74 Concluding remarks....159
7 Analogue tape reproduction....163
7.1 Preliminary remarks....163
7.2 Historical development of magnetic sound recording....164
7.3 Bias....166
7.4 Magnetised tape heads....168
7.5 Print-through....169
7.6 Azimuth....170
7.7 Frequency responses of tape recordings....172
7.8 "Standard" characteristics on open-reel tapes....173
7.9 Standards on tape cassettes....176
7.10 Operational principles....177
7.11 "Mono" and "half-track" tapes....180
7.12 "Twin-track" tapes....181
7.13 "Quarter-track" tapes....182
7.14 Practical reproduction issues....183
7.15 "Hi-fi" Tracks on domestic video....184
8 Optical film reproduction....187
8.1 Introduction....187
8.2 Optical sound with moving pictures....187
8.3 Considerations of strategy....188
8.4 Basic types of optical soundtracks....189
8.5 Soundtracks combined with optical picture media....190
8.6 Recovering the power-bandwidth product....191
8.7 Frequency responses....193

8.8 Reducing background noise 194
9 Reciprocal noise reduction 196
9.1 Principles of noise reduction 196
9.2 Non-reciprocal and reciprocal noise-reduction 196
9.3 Recognising reciprocal noise reduction systems 197
9.4 Principles of reciprocal noise reduction systems 198
9.5 Dolby "A" 199
9.6 Dolby "B" 201
9.7 DBX systems 202
9.8 JVC ANRS (Audio Noise Reduction System) 205
9.9 Telcom C4 205
9.10 Telefunken "High-Com" 206
9.11 The "CX" systems 206
9.12 Dolby "C" 207
9.13 Dolby SR and Dolby S 208
9.14 Other noise reduction systems 209
9.15 Noise reduction systems not needing treatment 211
9.16 Conclusion 211
10 Spatial recordings 213
10.1 Introduction 213
10.2 "Two-channel" recordings 213
10.3 Archaeological stereo 216
10.4 Matrixing into two channels 218
10.5 Three-channel recordings. 218
10.6 Four-channel recordings - in the cinema 219
10.7 Four-channel audio-only principles 220
10.8 Matrix "quadraphonic" systems 221
10.9 The QS system 221
10.10 The SQ system 222
10.11 Matrix H 223
10.12 "Dolby Stereo" 224
10.13 Developments of "Dolby Stereo" - (a) Dolby AC-3 226
10.14 Developments of "Dolby Stereo" - (b) Dolby headphone 227
10.15 Developments of "Dolby Stereo" - (c) Pro Logic 2 227
10.16 Discrete 4-channel systems - The JVC CD-4 system 227
10.17 The "UD-4" system 228
10.18 "Ambisonics" 229
10.19 Other discrete four-channel media 230
10.20 More than five-channel systems 230
10.21 Multitrack master-tapes 231
11 Dynamics 233
11.1 Introduction 233
11.2 The reasons for dynamic compression 234
11.3 Acoustic recording 235
11.4 Manually-controlled electrical recording 235
11.5 Procedures for reversing manual control 237
11.6 Automatic volume controlling 238
11.7 Principles of limiters and compressors 241
11.8 Identifying limited recordings 243
11.9 Attack times 244
11.10 Decay-times 245
11.11 The compression-ratio and how to kludge It 246
12 Acoustic recordings 249
12.1 Introduction 249
12.2 Ethical matters 249
12.3 Overall view of acoustic recording hardware 251
12.4 Performance modifications 252
12.5 Procedures for reverse-engineering the effects 254
12.6 Documentation of HMV acoustic recording equipment 255

12.7 The Recording horn - why horns were used 256
12.8 The lack of bass with horn recording 257
12.9 Resonances of air within the horn 257
12.10 Experimental methodology 257
12.11 Design of an Acoustic Horn Equaliser 259
12.12 Resonances of the horn itself 259
12.13 Positions of artists in relation to a conical horn 261
12.14 Acoustic impedances 262
12.15 Joining two horns 263
12.16 Joining three or more horns 264
12.17 Another way of connecting several horns 264
12.18 Electrical equalisation of recordings made with parallel-sided tubes 265
12.19 Electrical equalisation of recordings made with multiple recording horns 266
12.20 The recording soundbox 267
12.21 "Lumped" and "distributed" components 267
12.22 How pre-1925 soundboxes worked 268
12.23 Practicalities of acoustic recording diaphragms 270
12.24 The rest of the soundbox 271
12.25 Notes on variations 271
12.26 When we should apply these lessons 274
12.27 Summary of present-day equalisation possibilities 274
12.28 Conclusion 275
13 The engineer and the artist 279
13.1 Introduction 279
13.2 The effect of playing-time upon recorded performances 279
13.3 The introduction of the microphone 283
13.4 The performing environment 284
13.5 "Multitrack" issues 287
13.6 Signal strengths 289
13.7 Frequency ranges 290
13.8 Monitoring sound recordings 291
13.9 The effects of playback 292
13.10 The Costs of recording, making copies, and playback 295
13.10.1 The dawn of sound recording 295
13.10.2 Mass produced cylinders 297
13.10.3 Coarsegroove disc mastering costs 297
13.10.4 Retail prices of coarsegroove pressings 298
13.10.5 One-off disc recording 300
13.10.6 Magnetic tape costs 301
13.10.7 Pre-recorded tapes and cassettes 302
13.10.8 Popular music 303
13.10.9 Digital formats 303
13.10.10 Conclusion for pure sound recordings 304
13.11 The cinema and the performer 304
13.12 Film sound on disc 305
13.13 The art of film sound 307
13.14 Film sound editing and dubbing 309
13.15 The automatic volume limiter 311
13.16 Films, video, and the acoustic environment 312
13.17 Making continuous performances 313
13.18 Audio monitoring for visual media 313
13.19 Listening conditions and the target audience and its equipment 314
13.20 The influence of naturalism 316
Appendix 1. Preparing media for playback 318
Appendix 2: Aligning analogue tape reproducers 328

Preface

With the rapid pace of change in audio technology, analogue formats have all but disappeared as a means for current production and distribution of sound recordings. Nevertheless, many audio archivists are responsible for large numbers of valuable audio recordings in analogue formats. These require dedicated playback machines that have become obsolete, so the only way to ensure lasting access to this legacy is digitisation. To do this properly requires firstly that the optimum signal is extracted during playback from an analogue carrier, and this in turn necessitates extensive knowledge of the engineering processes and standards used at the time of its creation. The passing on of expertise and detailed knowledge gained during a time when analogue technology represented the cutting edge, is therefore of vital importance to subsequent generations, and it is with this in mind that this work was written.

The manual was written by Peter Copeland when he was Conservation Manager at the British Library Sound Archive over a ten-year period up until 2002, as an aid to audio engineers and audio archivists. Peter started his professional career at the BBC in 1961, initially working from Bush House in London for the World Service. From 1966 he worked as a dubbing engineer at BBC Bristol, taking on responsibility for sound in many productions by the famed Natural History Unit, among them the groundbreaking David Attenborough series, *Life on Earth*. In 1986 he joined the British Library as Conservation Manager for the Sound Archive where he started work on this manual. After Peter retired from the Library in 2002, we began work with him to polish the text ready for publication. The work was near to completion when Peter died in July 2006.

The original intention was to thoroughly edit the manual to bring it up to date, with appropriate diagrams and other illustrations added. However in the extremely fast-moving world of audiovisual archiving this would have entailed a great deal of rewriting, such that it would no longer have been the manuscript that Peter left us. After much discussion therefore, the decision has been taken to publish it now, largely unchanged from its original form, and make it freely available on the internet as a service to professional audiovisual engineers and archivists.

Readers should bear in mind that the manual was written over a period of time some years ago, and so should be considered as a snapshot of Peter's perspective during that time. Some information, particularly on digital conversion in chapter 3, is outdated: the R-DAT format is now obsolete, professional audio engineers routinely use computer-based hardware and software for audio processing with 24-bit converters at sample rates exceeding 96kHz, and sigma delta processors are now available. The core of the manual however, being concerned with making sense of the incredibly rich and complex history of analogue technology, remains a singular feat of rigorous, sustained research, and is unlikely to date. Some statements in the text represent personal opinion, although this is always clearly stated. The somewhat quirky, personal style in which the manual is written will be very familiar to anyone who knew Peter, and was as much a part of him as was his passion for the subject of this work. Minor amendments to the original texts have been made, mostly ironing out inconsistencies in style and for this I would like to thank Christine Adams, Nigel Bewley, Bill Lowry, Will Prentice, Andrew Pearson and Tom Ruane who helped check through the manual in its entirety.

This work assembles in one place a tremendous amount of factual information amassed during Peter's long career as an audio engineer - information that is difficult to find anywhere else. The breadth of research into the history of sound playback is unequalled. Peter himself sometimes referred to his role as that of an "audio archaeologist". This manual is a fitting and lasting testament to Peter's combination of depth of knowledge with clarity of expression.

Richard Ranft
The British Library Sound Archive
September 2008

Acknowledgements

Few types of torture can be more exquisite than writing a book about a subject which is instinctive to the writer. This has been the most difficult piece of writing I have ever undertaken, because everything lives in my brain in a very disorganised form. My first thanks must be to the inventors of the word processor, who allowed me to get my very thoughts into order.

Next I must acknowledge the many talented people who gave up considerable amounts of time to reading printouts at various stages. First, I would like to thank Lloyd Stickells, a former engineer of the British Library Sound Archive, who looked through a very early draft of this work while he was still with the Archive, and made sufficiently encouraging noises for me to want to continue. Steve Smolian, of Smolian Sound Studios, Potomac, Maryland, also made encouraging noises; moreover, he helped with many suggestions when I sought advice from the American point of view. Alistair Bamford also made encouraging noises as the work proceeded. Ted Kendall and Roger Wilmut read it more recently when it was nearly at full length, and Crispin Jewitt (Head of the Sound Archive) was rash enough to put his money where his mouth was, financing a Pilot Digitisation Project to test some of the methodology. As a result, Serena Lovelace was in the front line, and gave me very clear (and polite) feedback on every section of the book. But the biggest bouquet must go to my editor, David Way, who was obliged to read the whole thing many times over.

I must thank the following individuals for supplying me with specific pieces of information or actual artefacts, and I just hope I have reported the information correctly: The British Library Science Research and Information Service, Sean Davies (S. W. Davies Ltd.), Eliot Levin (Symposium Records), Alistair Murray, Noel Sidebottom (British Library Sound Archive), Pete Thomas (BBC), and Roger Wilmut. George Overstall, the guru of acoustic soundboxes and horns, gave me an afternoon of his time to the mechanics of treating and playing 78rpm discs; chapter 3 could not have been written without his help. Mrs. Ruth Edge, of the EMI Archive in Hayes, kindly allowed me to disassemble the world's last surviving professional acoustic recording machine, and measure its horns, without which Chapter 11 would have been impossible.

I was able to try several chapters of this manual upon an unsuspecting public, when Jack Wrigley asked me to write for his magazine *The Historic Record*. On the whole the readership did not argue with me, which boosted my morale despite my torture; the few comments I received have certainly resulted in greater clarity. So thanks to all Jack's readers as well.

But nothing at all could have been done without the continuous support of the Information Service of the British Library Sound Archive. It seems invidious to name anyone in particular, but I must thank Mavis Hammond-Brake, who happened to draw the short straw every time I required some information from a different department of the British Library. Sorry, Mavis, it wasn't planned that way! Since then, the same straw has been drawn by Lee Taylor - sorry, Lee.

Peter Copeland
London, 28th February 2001

1 Introduction

1.1 The organisation of this manual

This manual gives details of some of the techniques to be followed when old sound recordings are transferred to more modern carriers. It is aimed primarily at the professional archivist transferring them to conserve them, and it is generally assumed that the purpose is TO PRESERVE THE ORIGINAL SOUND. (That is what I understand to be the function of a "National *Sound* Archive", rather than preservation of the artefacts; that would be the function of a "Museum." Please feel free to disagree with me, though!)

Here is one disagreement I myself accept. In many cases the work of people behind the scenes is just as important as that of the performer, for example in editing defective sections of a performance; so this must often be modified to read "to preserve the original *intended* sound." I would enlarge this in two ways. When it comes to the subject matter, it must surely mean "intended by the *producer* of the recording" (or the film or the broadcast), although this will become rather a subjective judgement. And when it comes to technical matters, it must mean "Intended by the Sound Engineer." Hence this manual!

I also need to define the word "sound". Do we mean a psychoacoustic sensation, or objective variations in pressure at the ear(s)? In other words, when a tree fell in a prehistoric forest before animals evolved ears, did it make a "sound" or not? In this manual I use the second definition. It even seems possible that some form of genetic engineering may enable us to develop better ears (and brains to perceive the results, plus ways of storing and reproducing nerve pulses from the ears) in future. But the objective nature of sound pressures is what a sound archivist can (and must) preserve at present.

The arrangement of the book is as follows. After two chapters on overall copying strategy and the conversion of analogue sound to digital, we have five chapters on the techniques for getting the sound accurately from various analogue media. Each has some history and some scientific facts, and shows how we may use this knowledge to help us get back to the original sound today. The section on setting playing speeds, for example, covers both objective and subjective techniques, and contains a summary of our objective knowledge for reference purposes.

Next come three chapters for special techniques to ensure an old recording is heard the way the engineers originally intended. One deals with noise reduction systems, the second where spatial effects occur (e.g. stereo), and the third where the original dynamic range of a compressed recording might be recovered. These are all problems of *reproduction*, rather than of recording.

There are vast areas where we do not have objective knowledge, and we must rely upon future developments or discoveries. So I have left the discussion of acoustic recording techniques until this point. (I define "acoustic recordings" as "sound recordings made without the assistance of electronic amplification"). Besides our shortage of objective knowledge, the whole subject is much more complicated, and we must use previous vocabulary to express what little we know. Although large amounts of research are under way as I write, I can only indicate what I consider to be an appropriate strategy for a sound archive, which will give better fidelity for listeners until ideal technology becomes available.

In many other cases, we are comparatively close to our goal of restoring the original sound (or the original *intended* sound). We now have techniques former engineers could never have dreamt of, and even acoustic recordings will probably succumb to such progress. This brings three new difficulties.

In a recent broadcast ("Sunday Feature: Settling the Score" BBC Radio 3, Sunday 4th July 1999), the presenter Samuel West described how he had taken a (modern) performance of a solo piano piece, and deliberately distorted it until it sounded like an acoustic recording. He then played the two versions to a number of musically literate listeners. Not only were all the listeners fooled, but they put quite different artistic interpretations on their two responses, even though the actual performance was the same. This seems to show modern day listeners may have quite different artistic responses to historic records, if progress in sound restoration continues!

However, the programme then went on to outline the second of the three difficulties - the compromises forced upon performers by obsolete recording techniques. So my last chapter is an account of some techniques of former sound recording engineers, so you may judge how *they* balanced scientific and aesthetic thought processes, and understand some of the differences between the original sound and the intended original sound.

The third of the three difficulties is that sound recording is now becoming subservient to other media, because it is comparatively cheap and easy, and less of a self-contained discipline. Thus audio is becoming centred on applications, rather than technologies. All this means my final chapter will go out of date much faster than the rest of the book.

1.2 The target audience for this manual

Considerable research is needed to understand, and therefore to compensate for, the accidental distortions of old sound recording systems. This manual consists largely of the fruits of my research. Although the research has been extensive (and quite a lot of it is published for the first time), I do not intend this manual to fill the role of a formal historical paper. It is aimed at the operator with his hands on the controls, not the qualified engineer at his drawing board, nor the artist in his garret, nor the academic student of history.

So I have tried to describe technical matters in words, rather than using circuit diagrams or mathematical formulae. Some professional engineers will (correctly) criticise me for oversimplifying or using "subjective language." To these people, I plead guilty; but I hope the actual *results* of what I say will not be in error.

I have standardised upon some vocabulary which might otherwise cause confusion. In particular, what do we call the people who operated the original sound recording equipment, and what do we call the people restoring the sound today? I have adopted an arbitrary policy of calling the former "engineers" and the latter "operators." I apologise if this upsets or misrepresents anyone, but there is no universally accepted phraseology. The English language also lacks suitable pronouns for a single person who may be of either sex, so again I apologise if male pronouns suggest I am ignoring the ladies.

This manual includes a certain amount of sound recording history. I shall have to look at some aspects in great detail, because they have not been considered elsewhere; but to reduce unnecessary explanation, I have had to assume a certain level of historical knowledge on the part of the reader. For analogue media, that level is given by Roland

Gelatt's book *The Fabulous Phonograph* in its 1977 edition. If you want to get involved in reproducing the original sound and your knowledge of sound recording history isn't up to it, I strongly recommend you to digest that book first.

Psychoacoustics plays a large part, because recording engineers have intuitively used psychoacoustic tricks in their work. They have always been much easier to do than to describe, so the word "psychoacoustic" appears in most of my chapters! But it is very difficult to describe the tricks in scientific language. Furthermore, our present day knowledge is accumulated from long lines of scientific workers following in each other's footsteps - there are very few seminal papers in psychoacoustics, and new discoveries continue to be made. So I have not given references to such research, but I recommend another book if you're interested in this aspect: *An Introduction to the Psychology of Hearing b*y Brian C. Moore. However, you do not have to read that book *before* this one.

I must also explain that human beings are not *born* with the ability to hear. They have to *learn* it in the first eighteen months of their lives. For example, as they lie wriggling in their prams, Grandma might shake a rattle to get their attention. At first the child would not only be ignorant of the sound, but would lack the coordination of his other senses. Eventually he would turn his head and *see* the rattle, coordinating sight and sound to gain an understanding of what rattles are. There are six or seven senses being coordinated here, the sense of sight (which in this case is three senses combining to provide stereoscopic vision - the sense of left eye versus right eye, the sense of parallax, and the sense of the irises "pulling focus"), the sense of hearing (which is stereophonic, combining the difference in times and in amplitudes at the two ears), and the sense of balance and how this changes as the muscles of the neck operate. All this has to be *learnt*. Individual people learn in slightly different ways, and if an individual is defective in some physiological sense, psychological compensation may occur.

All this combines to make the sense of hearing remarkably complex. It is therefore even more amazing that, in the first 100 years of sound recording history, it was possible to fool the brain into thinking a sound recording was the real thing - and to a higher standard than any of the other senses.

A further difficulty I face is that of the reader's historical expertise. An expert can take one look at a disc record and immediately pronounce upon its age, rarity, what it will sound like, the surname of the recording engineer's mother-in-law, etc. An expert will be able to recognise the characteristics of a record just by looking at it. Much of my material will seem redundant to experts. The restoration operators employed by a single record company also do not need such detail, since they will be specialising in recordings whose characteristics are largely constant. But there are innumerable examples of operators getting it wrong when stepping beyond the areas they know. So I consider it important for every operator to read the book at least once, just to see how things may differ elsewhere.

The most difficult part is to know *which* technology was used for making a particular recording. This is practically impossible to teach. A recipe book approach with dates and numbers is easy to misunderstand, while the true expert relies on the "look and feel" of a particular artefact which is impossible to describe in words. I just hope that experts will not be upset by apparent trivia; but I have made a deliberate attempt to include such details if there is no convenient alternative. I must also confess that the archivist in me wants to get unwritten facts into print while it is still possible.

Yet another problem is caused by the frequent changes in hardware preferred by sound operators. So I shall not give recipe book instructions like "Use a Shure M44 cartridge for playing 78s," except when there are no alternatives. Instead I shall describe

the *principles*, and leave operators to implement them with the resources they have available. It is my opinion that transferring recordings made during the last century will continue for at least the next century. No-one can predict how the hardware will evolve in that time; but I am reasonably certain the principles will not change.

Past history also shows that you can sometimes have a "low-tech" and a "high-tech" solution to the same problem. Do not assume the "high-tech" solution will always give the best results. Practical skills - in handling and listening to old media - often outweigh the best that modern technology can offer. I can even think of some disgraceful cases where professional sound restoration operators have been thrown out of trade associations or engineering societies, because they favour "low-tech" solutions. Do not allow yourself to be corrupted by such ideas. There is always *one* optimum solution to a technical problem, and it is up to you to choose *that* solution. I cannot always teach you the craftsmanship aspects by means of the printed word; but I can, and do, explain the principles of recovering sound with optimum engineering quality. This will allow you to assess your own skills, and balance them against those of others, for yourself.

I cannot consider every sound medium which has ever been invented. I have therefore concentrated upon "mainstream" media, or media which illustrate a particular technical point I wish to make. More advanced or primitive types of sound recordings have been ignored, since when you know the basic principles, you will be able to figure out how to transfer them yourself. So you will not find the first stereo experiments mentioned, or early magnetic recordings, or freakish cylinders, or media never intended to have a long shelf life (such as those for dictation machines).

Finally, I shall only be describing the techniques I consider essential for someone trying to restore the original intended sound. There are many others. People have said to me "Why don't you mention such and such?" It is usually because I disapprove of "such and such" on principle. I shall be placing great emphasis upon ways giving the most accurate results, and I shall simply ignore the rest - otherwise the length of the book would be doubled.

1.3 The original sound

I had been a professional sound engineer for a quarter of a century before I joined the British Library Sound Archive. I was the first such person to be appointed as a "Conservation Manager," in overall charge of sound conservation strategy. And I must make it clear that I was regarded as an outsider by the library community. ("A *sound* expert? Working in a *Library*!?")

As soon as I had cleared away the short ends of quarter-inch tape which stopped me putting my feet under my desk, I soon became aware that the "culture" of the sound engineer was not appreciated within the building. So I ask your forgiveness if this manual appears to say "You must do this" or "You should do that" with no rational explanation being given. The culture of a successful analogue sound operator appears to be learnt at his mother's knee. It isn't always clear whether he's generated the rules himself, learnt them from his peers, or had it hammered into him on a formal training course. To illustrate my meaning, the very first task I witnessed my staff doing was a straight copy of a quarter-inch analogue tape. The operator never even looked at the stroboscope of the playback deck to check if it was running at the right speed! However, he *did* go through a prescribed procedure for checking the performance of the destination machine with tone signals; but he obviously did not understand *why* this was being done, he only knew who

to ask if he got the wrong figures on the meter. All this would be second nature to a professional sound operator from his mother's knee. So I apologise again if I keep stating what I consider to be obvious.

A broader problem is this. I am the first to recognise that "restoring the original intended sound" may not be the motivation for all transfer operators. The success of Robert Parker in making old recordings accessible to the modern public is proof of that, and he has been followed by numerous other workers bringing various corners of the recorded repertoire out of obscurity. Parker has been the subject of criticism for imposing artificial reverberation and fake stereo effects upon the original transfers. He replies that without these techniques the music would not have had such a wide appeal, and anyway he has lodged the untreated tapes in his vault. I think this is the right attitude. Even if commercial release is anticipated, I consider that recovering the "original sound" should always be the first step, whatever happens to it afterwards. Describing subjective improvements would again double the length of this manual and cause it to go out of date very rapidly (partly because of changes in fashion, and partly because of new technical processes). But I hope my remarks will also be helpful when *exploitation* is the driving force, rather than preservation.

Since older media often distorted the sound, it is first necessary to decide *whether* we should attempt to restore the sound in an undistorted form. It is often argued that the existing media should be transferred as they are, warts and all, on the grounds that better restoration technology may be available in the future. Another argument says that such warts are part of the ambiance in which such media were appreciated in the past, and should be preserved as a significant part of the artefact.

Having been a professional recording engineer myself, I challenge these views. I should not wish that the sound recordings I made before I joined the British Library Sound Archive should be reproduced "warts and all". I should certainly demand that the ravages of time, and the undocumented but deliberate distortions (the "recording characteristics"), should always be compensated, because listeners will then get my original intended sound. So I consider it's my responsibility to perform similar services for my predecessors. As for attempts to tidy up my work in ways which weren't possible when I made the recordings, I hold the view that where the warts are accidental (as opposed to deliberate distortions, such as might be applied to a guitar within a pop music balance), I have no objection to their being corrected, so long as the corrections result in more faithful intended sound.

I shall now respond to the assumption of the "warts and all brigade" that future technology will be better than ours. Frankly, I am not fully convinced by this argument, because with a scientific approach we can usually quantify the effects of technology, and decide whether or not future technology can offer any improvement. I only fear technology when it doesn't exist at all, or when it exists in the form of "trade secrets" which I cannot judge. (I shall be indicating these cases as we come to them). Rather, I fear that sound recording will become more and more "idiot-proof," and eventually we shall forget the relationships between past artists and engineers. If we misunderstand this relationship, we are likely to misunderstand the way the recording equipment was used, and we will be unable to reproduce the sounds correctly, even with perfect technology. I shall illustrate the point with the same example I mentioned above. Enjoying popular music when I was young, I generally know which distortions were deliberate - the guitar in the pop mix - and I know which were accidental; but I must not assume these points will always be appreciated in the future. Indeed, I strongly suspect that the passage of time will make it more difficult for future operators to appreciate what is now subliminal

for us. But few people appreciate these "cultural" factors. They have never been written down; but they're there, so I shall be making some references to them in the final chapter.

I shall, however, mention one now. Recording sound to accompany pictures is a completely different business from recording sound on its own. I have spent much of my life as a film and video dubbing mixer, and I cannot think of a single case where it would be justifiable to take any of my final mixes and "restore the original sound," even if it were possible. I would only want people to go as far as indicated above - to undo the ravages of time and equalise the known recording characteristics. All the rest of the "distortions" are deliberate - to distract from compromises made during the picture shooting process, to steer the emotional direction of the film by the addition of music and/or the pace of the mixing, to deliberately drive the dynamics of the sound to fit imperfect pictures, etc. In these circumstances pictures are dominant while sound is subservient - the sound only *helps* to convey a film's message. (Films of musical performances seem to be the principal exception).

Most people find their visual sense is stronger than their aural sense, even though sound recording has achieved a higher degree of "fidelity" than moving pictures. Thus films and videos become art-forms with rules of their own, built into them at the time they went through "post-production." When we *do* want to restore the "original sound," rather than the original *intended* sound, we should clearly divorce the sound from the pictures, and use "rushes" or other raw material in an unmixed state rather than the final mix.

Finally, I should like to mention that some workers have argued that old recordings should be played on old equipment, so we would hear them the way contemporary engineers intended. I have a certain amount of sympathy with this view, although it does not agree with my own opinion. I would prefer my recordings to be played on state-of-the-art equipment, not what I had thirty years ago! But if we wish to pursue this avenue, we meet other difficulties. The principal one is that we have very few accounts of the hardware actually used by contemporary engineers, so we don't actually know what is "right" for the way they worked.

Even if we did have this knowledge, we would have to maintain the preserved equipment to contemporary standards. There was a great deal of craftsmanship and taste involved in this, which cannot be maintained by recipe book methods. Next we would need an enormous collection of such equipment, possibly one piece for every half decade and every format, to satisfy any legitimate historical demand for sound the way the original workers heard it. And we would inevitably cause a lot of wear and tear to our collection of original recordings, as we do not have satisfactory ways of making modern replicas of original records.

But it so happens that we can have our cake and eat it. If we transfer the sound electrically using precise objective techniques, we can recreate the sound of that record being played on *any* reproducing machine at a subsequent date. For example, we could drive its amplifier from our replayed copy, its soundbox from a motional feedback transducer, or its aerial from an RF modulator.

1.4 Operational principles

I shall first state what I believe to be an extremely important principle. I believe the goal of the present-day restoration expert should be to compensate for the known deficiencies *objectively*. He should not start by playing the recording and twiddling the

knobs *subjectively*. He should have the courtesy first to reproduce the sound with all known *objective* parameters compensated. For archival purposes, this could be the end of the matter; but it may happen that some minor deficiencies remain which were not apparent (or curable) to contemporary engineers, and these can next be undone. In any event, I personally think that *only* when the known objective parameters have been compensated does *anyone* have the moral right to fiddle subjectively - whether in an archive, or for exploitation.

The aim of objectivity implies that we should *measure* what we are doing. In fact, considerable engineering work may be needed to ensure all the apparatus is performing to specification. I know this goes against the grain for some people, who take the view that "the ear should be the final arbiter." My view is that of course the ear should be the *final* arbiter. But, even as a professional recording engineer deeply concerned with artistic effects, I maintain that measurements should come first. "Understanding comes from measurement" as physical scientists say; if we can measure something's wrong, then clearly it *is* wrong.

On numerous occasions, history has shown that listeners have perceived something wrong before the techniques for measuring it were developed; this is bound to continue. Unfortunately, "golden-eared" listeners are frequently people who are technically illiterate, unable to describe the problem in terms an engineer would understand. My personal view (which you are always free to reject if you wish), is that measurements come first; then proper statistically-based double-blind trials with "golden-eared" listeners to establish there *is* a valid basis for complaining about problems; then only when this has been done can we reasonably research ways to cure the problem. I certainly do not wish to discourage you from careful listening; but accurate sound reproduction must at the very least begin with equipment whose performance *measures* correctly.

On the other hand, the ear is also important in a rather coarse sense - to get us back on the right track if we are catastrophically wrong. For example, if the tape box label says the tape runs at 15 inches per second and the tape sounds as if it's at double speed, then it will probably be a fault in the documentation, not a fault in our ears!

For intermediate cases, we should be able to justify subjective decisions in objective terms. For example, if we switch the tape reproducer to 7.5 inches per second and we perceive music at slightly the wrong pitch, then we should proceed as follows. First we check our own sense of pitch with a known frequency source properly calibrated. Then we quantify the error and we seek explanations. (Was it an unreliable tape recorder? or an historic musical instrument?) If we cannot find an explanation, we then seek confirmatory evidence. (Is the background hum similarly pitch-shifted? Does the tape play for the correct duration?) But, at the end of the day, if there is no objective explanation, a sound archive *must* transfer the tape so that at least one copy is exactly like the original, regardless of the evidence of our senses.

The question then arises, which subjective compensations should be done in the environment of a sound archive? A strictly scientific approach might suggest that no such compensations should ever be considered. But most professional audio operators are recruited from a background which includes both the arts and the sciences. It is my personal belief that this is only to the good, because if these elements are correctly balanced, one doesn't dominate over the other. But it is impossible for anyone's artistic expertise to stretch across the whole range of recorded sound. It may be necessary to restrict the artistic involvement of an operator, depending upon the breadth of his

knowledge. To be brutal about it, a pop expert may know about the deliberately distorted guitar, whereas an expert in languages may not.

This assertion has not answered the potential criticism that the artistic input should ideally be zero. I shall counter that argument by way of an example. For the past thirty years Britain has had an expert in restoring jazz and certain types of dance music in the person of John R. T. Davies. I do not think he will mind if I say that his scientific knowledge is not particularly great; but as he played in the Temperance Seven, and has collected old records of his own particular genre ever since he was a boy, his knowledge of what such music *should* sound like has been his greatest asset. Long before the present scientific knowledge came to be formulated, he had deduced it *by ear.* He therefore systematically acquired the hardware he needed to eliminate the various distortions he perceived, and although the methods he evolved appear a little odd to someone like me with a scientific background, he ends up with very accurate results. John Davies does not claim to be perfect, but he holds the position that his particular musical knowledge prevents him making silly mistakes of the type which might befall a zombie operator.

I therefore accept that a certain level of artistic input is advantageous, if only to insure against some of the howlers of a zombie operator. But my position on the matter is that each individual must recognise his own limited knowledge, and never to go beyond it. We shall be encountering cases in the following pages where specialist artistic knowledge is vital. When such knowledge comes from an artistic expert, I consider it is no less reliable than pure scientific knowledge; but I would feel entitled to query it if I wasn't certain the knowledge was "right". Such knowledge may not just be about the original performance. It may also be knowledge of the relationship between the artist and the contemporary recording engineer. It is not always realised that the sounds may have been modified as they were being created, for very good reasons at the time.

1.5 A quadruple conservation strategy

The contradictions which can arise between "technical" and cultural" factors have caused passionate debate within the British Library Sound Archive. Indeed, the writer once addressed an external public meeting on the subject, and the audience nearly came to blows. There is often no satisfactory compromise which can be taken between the contradictions. I propose to leave some of the considerations until we get to the final chapter; but in the meantime the best answer seems to be a "quadruple conservation" strategy. This would mean the archive might end up with as many as four versions of a recording for long term storage, although two or more might often be combined into one.

(1) The original, kept for as long as it lasts.
(2) A copy with warts-and-all, sounding as much like the original artefact as possible, which I shall call "The Archive Copy."
(3) A copy with all known objective parameters compensated, which I shall call "The Objective Copy."
(4) A copy with all known subjective and cultural parameters compensated, which I shall call "The Service Copy."

I recognise that such an ambitious policy may not always be possible. As we reach critical decision points during the course of this manual, I shall be giving my personal recommendations; but I am writing from the point of view of a professional sound

recordist in a publicly funded national archive. Each reader must make his own decision for himself (or his employer) once he understands the issues.

Fortunately, there are also many occasions where, even from an ivory tower viewpoint, we don't actually need an original plus three copies. For example, when basic engineering principles tell us we have recovered the sound as well as we can, the objective and service copies might well be identical. Sometimes we do not have the knowledge to do an "objective copy", and sometimes cultural pressures are so intense that we might never do an "archive copy." (I shall describe an example of the latter in section 13.2). But I shall assume the quadruple conservation strategy lies behind our various attempts, and I advise my readers to remember these names as the work proceeds.

1.6 How to achieve objectivity

Given that the purpose of conservation copying is to "restore the original intended sound", how do we go about this? How can we know we are doing the job with as much objectivity as possible, especially with older media made with temperamental recording machinery, or before the days of international standards? The present-day archivist has the following sources of knowledge to help him, which I list in approximate order of importance.

1. Contemporary "objective recordings." This generally means contemporary frequency discs or tapes, together with a small handful of other engineering test media for intermodulation distortion and speed. Many large manufacturers made test recordings, if only for testing the reproducers they made, and provided they have unambiguous written documentation, we can use them to calibrate modern reproducers to give the correct result. Unfortunately, not all test records are unambiguously documented, and I shall allude to such cases as they arise. Even worse, many manufacturers did not make test recordings. Yet objective measurements are sometimes available accidentally. For instance, several workers have analysed the "white noise" of the swarf vacuum pipe near an acoustic recording diaphragm to establish the diaphragm's resonant frequency. And we will be coming across another rather curious example in section 6.47.
2. Contemporary, or near-contemporary, objective measurements of the recording equipment. Written accounts of contemporary measurements are preferable to present-day measurements, because it is more likely the machinery was set up using contemporary methods of alignment, undecayed rubber, new valves, magnetic materials at the appropriate strength, etc. As for the measuring equipment, there are practically no cases where present-day test gear would give significantly different results from contemporary measuring equipment. As the machinery improved, so did measuring equipment; so contemporary measurements will always be of the right order of magnitude.

On the other hand, alleged objective measurements made by the equipment makers themselves should always be subject to deep suspicion (this applies particularly to microphones). This last point reminds us that there may be "specifications" for old recording equipment, besides "objective measurements." These must be regarded with even more suspicion; but when there were international standards for sound recordings, we must at least examine how well the equipment actually conformed to the standards.

3. Present-day analysis of surviving equipment. This suffers from the disadvantages hinted at above, where the equipment isn't necessarily in its best state. There are also the "cultural" factors; the way in which the machinery was used was often very important, and may invalidate any scientific results achieved.
4. Analysis of drawings, patents, etc. It is often possible to work out the performance of a piece of recording equipment from documentary evidence; no actual artefact or replica is required.
5. Interviews with the engineers concerned. Reminiscences of the people who operated old recording equipment often reveal objective information (and sometimes trade secrets).
6. Reverse engineering surviving recordings. In general, this is not possible for one recording alone; but if a number of different recordings made by the same organisation exhibit similar characteristics, it is possible to assume they are characteristics of the machine which made them, rather than of the performances. It is therefore possible to reverse engineer the machine and neutralise the characteristics. A similar technique occurs when we have a recording which is demonstrably a copy of another. We can then use the original to deduce the characteristics of the copy, and thence other recordings made on that equipment.
7. Automatic analysis. It is early days yet, but I mention this because mathematical analysis of digital transfers is being invoked to identify resonances in the original. One aim is to eliminate the "tinnyness" of acoustic horn recording. Identifying resonances by such an objective technique is clearly superior to the subjective approach of the next suggestion.
8. Intelligent artistic input. If it's known that a type of recording had resonances, it may be possible to listen out for them and neutralise them by ear. This implies there is a "general structure" to the recorded characteristics, which can be neutralised by appropriate "tuning." So in the following pages I have included a few such "general structures." But I've placed this evidence last, because it's very easy to cross the borderline between objective and subjective compensation. Although the result may be more faithful than no compensation at all, there will be no proof that the particular form of tuning is exactly correct, and this may give difficulties to subsequent generations who inherit our copies. As digital signal processing evolves, it will be feasible to do statistical analysis to determine a "level of confidence" in the results. Then, at some point (which would need a consensus among archivists), the process might even be adopted for the "objective copy".

1.7 The necessity for documentation

If we continue to assume that "recovering the intended original sound" is a vital stage of the process, this implies that we should always keep some documentation of what we have done. It might take the form of a written "recording report" or a spoken announcement. It has three functions. It advises our successors of any subjective elements in our work, enabling them to reverse engineer it if the need arises after the original has decayed away. It also shows our successors the steps we have taken ourselves and our thought processes, so later generations will not be tempted to undo or repeat our work without good reason. I must also say that I personally find the documentation useful for a third reason, although not everyone will agree with me. I find it forces me to think each case through logically.

2 The overall copying strategy

2.1 The problem to be solved

In this manual I do not propose to discuss the major strategies of running a sound archive; instead, I shall refer you to a book by my mentor Alan Ward (*A Manual of Sound Archive Administration*, pub. Gower, 1990). But this chapter includes wider issues than just analogue sound reproduction and copying.

Some philosophers have considered the possibility of a replicating machine which might build an exact replica of an original recording, atom by atom. This is science fiction at present, so the only other way is to play such a recording back and re-record it.

But even if we could build such a replicating machine, I suspect that the universe may contain something more fundamental even than sub-atomic particles. Here is a rhetorical question for you to ponder: What is "Information"?

It may even be what holds the Universe together! When certain sub-atomic particles separate under the laws of Quantum Physics, they may be connected by "information" which travels even faster than light, but which does not actually travel until you make the observation. This is still a novel concept amongst the scientific community as I write (Ref. 1); but within a few decades I suspect it will be as familiar to schoolchildren as "Relativity" is now. And, since sound recording is by definition a way of storing "information," such philosophical issues aren't completely irrelevant to us.

2.2 General issues

Most of this manual is designed to facilitate the playback process so as to recover the information - the sound - without any intentional or unintentional distortions. It is aimed at the operator whose hands are on the controls, rather than the manager planning the overall strategy. For the latter, politics, cost, space and time are paramount; he is less concerned with "mechanics." But it would be wrong for me to ignore the operational aspects of overall strategy, if only because in smaller archives the manager and the operator is the same person; so I shall now say a few words on the subject.

First, the law of copyright. This differs from one country to the next, and may also have exemptions for archival applications. For many years the British Library Sound Archive had special permission from The British Phonographic Industry Ltd. to make copies of records for internal purposes, since published records had no "fair dealing" exemptions. Under procedures laid down under the 1988 Copyright Act, archival copying work might always then be possible provided the Secretary for State was persuaded that the archive was "not conducted principally for profit"; but I must stress that, whatever I recommend, it does not absolve you from observing the law of copyright in your country.

The manager will certainly be concerned with cost, perhaps thinking of getting the maximum amount of work done for a particular budget. Frankly, I believe this is inappropriate for an archive dedicated to conserving sounds for centuries, but I recognise this will be a consideration in the commercial world. A manager must therefore understand the principles, so he may see clearly how the work will suffer if the ideal scenario is not followed. It may not be a catastrophe if it isn't, but there will be trade-offs. The procedure actually used should certainly be documented, and then originals should be kept so that future generations can have another bite at the cherry. So the manager must

assess the costs of storing the originals and then financing another bite of the cherry, comparing them with the costs of the ideal scenario.

I am afraid that experience also shows that "unexpected hitches" are frequent. It is usually impossible to copy sounds using production-line techniques. Whatever overall strategy you adopt, your schedule is certain to be disrupted sooner or later by a recording which requires many times the man-hours of apparently-similar items.

2.3 The principle of the "Power-Bandwidth Product"

As I said, the only way of conserving sounds which are at risk is to copy them. "At risk" can mean theft, wilful destruction, accidental erasure, biological attack, miscataloguing, or wear-and-tear, as well as plain chemical breakdown. But if it's considered there is little chance of these, then there is much to be said for simply keeping the original recording uncopied, for the following reason.

Analogue recordings cannot be copied without some loss of quality, or "information" as engineers call it. Despite the idea of "information" being a fundamental property of matter, to an analogue engineer "information" is an objective measurement of the quality of a recording. It is obtained by multiplying the frequency range, by the number of decibels between the power of the loudest undistorted signal and the power of the background noise. The result is the "power-bandwidth product." This term is used by analogue engineers to measure the information-carrying capacity of such things as transformers, landlines, and satellites, besides sound recordings.

It is always possible to trade one parameter against the other. To return to sound recording, a hissy disc may have a full frequency range to the limits of human hearing (say 16kHz), but if we apply a high-frequency filter when we play it, the hiss is reduced. In fact, if the filter is set to 8kHz, so that the frequency range is halved, the hiss will also be halved in power. We can therefore trade frequency range against background noise. Of course, there may be other parameters which are not covered by the power-bandwidth formula - such as speed constancy - but because it's a fundamental limitation, we must always consider it first. It is true that in Chapter 3 we may learn about a potential process for breaching the background-noise barrier *without* touching the wanted sound; but that process is not yet available, and in any case we must always consider the matter "downstream" of us. In objective terms, there's no way round it. (For further details, see Box 2.3).

The first strategic point about copying analogue sound recordings is therefore to minimise the loss of power-bandwidth product caused by the copying process. If the hissy disc mentioned above were copied to *another* hissy disc with the same performance, the hiss would be doubled, and we would irrevocably lose half the power-bandwidth product of the original. An archive should therefore copy analogue recordings to another medium which has a much greater power-bandwidth product, to minimise the inherent losses.

BOX 2.3
THE POWER-BANDWIDTH PRODUCT IN AUDIO RECORDINGS

This box is aimed at engineers. It is relatively easy to assess the information-carrying capacity of analogue devices such as transformers, landlines, and satellites. They tend to have flat responses within the passband, Gaussian noise characteristics, and clear overload points. But sound recordings generally do not have these features, so I must explain how we might quantify the power-bandwidth product of sound recordings.

Analogue media overload "gently" - the distortion gradually gets worse as the signal volume increases. So we must make an arbitrary definition of "overload." In professional analogue audio circles, two percent total harmonic distortion was generally assumed. As this is realistic for most of the analogue media we shall be considering, I propose to stick to this.

For electronic devices, the bandwidth is conventionally assumed to be the points where the frequency response has fallen to half-power. This is distinctly misleading for sound recordings, which often have very uneven responses; the unevenness frequently exceeds a factor of two. There is another complication as well (see section 2.4). For my purposes, I propose to alter my definition of "bandwidth" to mean the point at which the signal is *equal* to random (Gaussian) noise - a much wider definition. Yet this is not unrealistic, because random noise is in principle unpredictable, so we can never neutralise it. We can only circumvent it by relying upon psychoacoustics or, for particular combinations of circumstances (as we shall see in Chapter 3). Thus random noise tends to form the baseline beyond which we cannot go without introducing subjectivism, so this definition has the advantage that it also forms the limit to what is objectively possible.

But most recording media do not have Gaussian noise characteristics. After we have eliminated the predictable components of noise, even their random noise varies with frequency in a non-Gaussian way. We must perform a spectral analysis of the medium to quantify how the noise varies with frequency. And because we can (in principle) equalise frequency-response errors (causing an analogous alteration to the noise spectrum), the *difference* between the recorded frequency-response and the noise-spectrum is what we should measure.

The human ear's perception of both frequencies and sound-power is a "logarithmic" one. Thus, every time a frequency is doubled, the interval sounds the same (an "octave"), and every time the sound power increases by three decibels the subjective effect of the increase is also very similar to other three-decibel increases. Following the way analogue sound engineers work, my assessment of the power-bandwidth product of an analogue sound recording is therefore to plot the frequency response at the 2% harmonic-distortion level, and the noise spectrum, on a log-log graph; and measure the AREA between the two curves. The bigger the area, the more information the recording holds.

2.4 Restricting the bandwidth

With an older record, we may be tempted to say "There's nothing above 8 kiloHertz, so we can copy it with the hiss filtered off without losing any of the wanted sound, and make it more pleasant to listen to at the same time." This is a very common attitude, and I want to take some space to demolish the idea, because it is definitely wrong for archival copying, although it might be justifiable for exploitation work.

The first point is that if it ever becomes possible to restore the frequencies *above* 8kHz somehow, three considerations will make it more difficult for our successors. First, the copy will add its own hiss above 8kHz, perhaps much fainter than that of the original; but when the original high frequencies are further attenuated by the filter, the wanted signal will be drowned more efficiently. Secondly, by making the high frequencies weaker, we shall make it much more difficult for our successors to assess and pursue what little there *is*. Thirdly, we actually have means for restoring some of the missing frequencies *now* - imperfectly and subjectively, it is true; but to eliminate such sounds with filters is an act of destruction exactly analogous to theft, erasure, or wear-and-tear.

The second point is, how do we "know" the "fact" that there is "nothing" above 8 kiloHertz? Actually, there is no known method for cutting all sounds off above (or below) a fixed frequency. Whether the effect is done acoustically, mechanically, or electronically, all such systems have a *slope* in their frequency responses. A disc-recording cutterhead, for example, may work up to 8kHz, and above that its response will slope away at twelve decibels per octave, so that at 16kHz the cutter will be recording twelve decibels less efficiently. So it is never true to say there's *nothing* above 8kHz. In a well-matched system, the performance of the microphone, the amplifier, and the cutterhead will be very similar, so the overall result might be a slope of as much as 36 decibels per octave; but this hardly ever seems to happen. Certainly, experiments have shown that there *is* audible information above the "official" limit. Often it is highly distorted and difficult to amplify without blowing up the loudspeaker with hiss, but it's there all right. The question then becomes "how do we go about making the high frequencies more audible", rather than "where do we cut them off."

I regret having to labour this point, but a very respected digital sound expert once fell into this trap. He did a computer analysis of the energy of an acoustic recording at different frequencies, and observed that noise dominated above 4kHz, so ruthlessly cut those frequencies off. In his published paper he devoted some puzzled paragraphs to why experienced listeners found the resulting recordings muffled. The reason, of course, is that (using psychoacoustics) human beings can hear many sounds when they are twenty or thirty decibels *fainter* than noise.

The only possible justification for filtering is if the subsequent recording medium is about to overload. Fortunately digital media are relatively immune from such problems, but it is a perpetual problem with analogue copy media.

In practice, this recovery of sound above an official cut-off frequency is a technique in its infancy. We can do it, but as Peter Eckersley is reported to have said, "the wider you open a window, the more muck blows in." Practical "muck" comprises both background-noise and distortion-products. The techniques for removing these are in their infancy. Unless such sounds can be restored perfectly, it is probably better that we should not try. But it is equally wrong for us to throw them away. The logical compromise is to transfer the high frequencies "flat" on the "archive copy," so future researchers will have

the raw material to work on. The copy medium must therefore have suitable power-bandwidth characteristics so that it will not alter the noise or distortion of the original medium. From the present state-of-the-art, we suspect that harmonic-distortion removal will depend critically upon phase linearity; therefore the copy medium must not introduce phase distortion either, or if it does it must be documented somehow (e. g. by recording its impulse-response - see section 3.4).

2.5 Deciding priorities

Our strategy is therefore to copy the vulnerable recording to another medium which has ample power-bandwidth product so that we don't lose very much, and not to filter the recording. Unfortunately, all media have a finite power-bandwidth product, so in fact we shall always lose *something*. The strategy must therefore balance the inevitable losses against the financial costs of making the copy and the likelihood of the original surviving to another day. This adds another dimension to the equation, because when analogue media degrade, their power-bandwidth product suffers (it's usually because their background noise goes up). So, one must decide *when* to do one's copying programme, depending on the power-bandwidth product of the original, its likely power-bandwidth product after future years in storage, the ability to recover power-bandwidth product at any particular point in time, and the power-bandwidth capacity of the copy medium.

Clearly we must always give top priority to media whose power-bandwidth product seems to be degrading faster than we can reproduce it. At the British Library this means wax cylinders and cellulose nitrate discs. (Acetate tapes are also vulnerable because the base-material is getting more brittle, but this does not directly affect the power-bandwidth product). There is a race against time to save these media, and the matter is not helped by two further circumstances. These media tend to be less-well documented, and it is impossible to play them without risk of damage; so the sounds must be copied *before* anyone can make an informed judgement on whether it is worth copying them ! Other media are less vulnerable, so we can afford to make a considered judgement about when to start copying them, and the balance will tilt as our knowledge improves. Also it is quite possible (although, in this digital age, less likely) that the technology for obtaining the maximum power-bandwidth product will improve. I am not going to talk about the present state-of-the-art here, because any such discussion will quickly go out of date; but I believe the basic principles will not change.

2.6 Getting the best original power-bandwidth product

The power-bandwidth product of an analogue recording always suffers if it is copied, so we must ensure we are working with "an original", not a copy. Thus we need to know the provenance of the analogue record. Does an earlier generation exist elsewhere? Does a manufacturer have master-tapes or metal negatives in his vault? Do masters exist of copies donated to your archive? We find ourselves picking up political hot potatoes when we examine this aspect, but the issue must be faced.

A knowledge of sound recording history is vital here, or at least that part of sound recording history which has a bearing upon your particular archive. The ideal strategy would be to collect full lists of the holdings of originals in various collections; but an adequate substitute might take the form of a generalised statement. At the British Library

Sound Archive, we have an interest in commercial records made by Britain's leading manufacturer EMI Records Ltd. It is useful for us to know that: "The British EMI factory has disposed of the metalwork for all recordings of black-label status or below, which were deleted by the outbreak of the Second World War." This sentence shows us the recordings we must seek and process in order to complete the collection for the nation. I advise you to collect similar statements to describe the genres you are interested in.

The strategy will also be determined by which media have the greatest power-bandwidth product, not just their mere existence. Although the metalwork mentioned in the previous paragraph is amongst the most rugged of all sound recording media, that situation isn't always the case. From about 1940 onwards, for example, Columbia Records in the United States did their mastering on large nitrate discs in anticipation of long-playing records ("L.P."s), which they introduced in 1948. These nitrates, if they still survive today, will be nearing the end of their useful life. Similar considerations apply to early tape. Thus a properly-planned conservation strategy will also take account of the lifetime of the "masters."

The archive must have some idea about the existence or non-existence of such holdings, because I frankly don't see the point of wasting time recovering the original sound from a second or third-generation copy when a version with a better power-bandwidth product exists somewhere else. The only reason might be to make service copies for use as long as the earlier generation remains inaccessible, or "partially objective" copies for reference purposes in the future (I shall explain this idea in section 2.8).

The overall strategy should be planned in such a way that, if a better version turns up, you can substitute it. There should be minimum disruption despite the obvious difficulties. Re-cataloguing must be possible to ensure the old version doesn't get used by mistake, but documentation of the old one should not be destroyed.

With nearly all analogue media it is easy to establish the order of the generations by ear - exact provenances aren't always essential. For example, if an analogue tape is copied onto a similar machine similarly aligned, the hiss will double, the wow-and-flutter will double, and the distortion will double. These effects are quite easy to hear so long as the two tapes are running simultaneously into a changeover switch under the control of the operator. Difficulties only occur when the two tapes are on opposite sides of the world and neither owner will allow them to be moved, or one is suspected of being a copy of the other on a medium with a better power-bandwidth product (but it isn't certain which is the original and which is the copy), or the original has disappeared and you must choose between two different copies of the same generation.

This latter case, two or more copies of an "original," is not uncommon. If the original has disappeared, it behoves us to choose the copy with the maximum power-bandwidth product. To put it more simply, if there are two copies available, we must choose the better one. It seems almost self-evident; but it's a principle which is often ignored.

A further dimension is that it may be possible to combine two copies to get an even better power-bandwidth product than either of them alone, and we shall be looking at this in Chapter 3. There may be political and practical difficulties; but every effort should be put into securing several good copies *before* the copying session starts.

Copies manufactured in other countries may often be better quality than locally-made ones. Meanwhile, recordings made *for* foreign broadcasters may only survive in *foreign* vaults. Thus you may be kept very busy chasing versions in other countries, usually with different catalogue numbers.

All this means that someone qualified to do discographical work may be kept just as busy as the actual sound operator. The two should work in close collaboration for another reason as well. Often technical factors depend on the date of the recording, or its publication-date; so the operator (or manager) may need this information before work starts.

2.7 Archive, objective, and service copies

With these considerations in mind, it now seems appropriate to address the issue of the versions we wish to make. (We considered the "three possible copies" in section 1.5) Until now, most copying has been "demand-led" - the demand from listeners and customers dictates what gets copied. While this is all right so far as it goes, the result is usually that only "service copies" are achieved, because copies are tailored to listeners' needs with subjective and cultural factors incorporated.

In my view, a proper programme of archival copying cannot be demand-led for that reason, and the following as well. The technical standards for service copies can be less critical, so general standards are lowered; I confess I have been guilty of this myself. Service copies are often done "against the clock", when loving care-and-attention is in short supply. And since the demand always comes from someone familiar with the subject matter, documentation tends to be less rigorously done.

Thus a programme incorporating several separate copies will take longer as well. It may be necessary to do three versions and document them. And it is advisable to have a procedure to prevent the same job being done twice.

On the other hand, there are ways to save time if a proper programme is planned. Demand-led hopping between different media with different characteristics wastes time connecting and aligning equipment, and may mean research and experiment if the plan does not confine itself to known areas. It requires "technical rehearsal time," which I shall consider shortly. Thus it is best to allocate at least a full working day specifically to archival copying without risk of interruption, and during that time a slab of technically-similar technically-understood work should be tackled.

There are many cases in which the various copy versions may be combined. If a disc record is so good that modern technology can do nothing to improve the sound, then the objective and service copies might as well be identical. Many professionally-made tapes can be copied to fill all three roles.

The overall strategy must always be capable of giving predictable results. If two different operators do the same job with different equipment, there should be no audible difference between their two "archive copies" and their two "objective copies". This implies that the operators should be supported by technical staff ensuring that all the equipment operates to international standards. A programme of routine measurement of equipment is essential, and if a machine is discovered to have been operated in a misaligned state, all the work done by that machine in the meantime should be checked and, if necessary, re-done. I shall not impose my ideas of the tolerances needed in such measurements, as standards are bound to rise with time; but managers must ensure such checks take place at frequent intervals.

Top-of-the-range copying facilities have high capital costs. These might be diluted by arranging a shift-system, so the equipment is in constant use. Alternatively, one shift might be doing exploitation work while another is doing strict archival work and a third is doing routine maintenance.

2.8 "Partially objective" copies

Sometimes the maximum power-bandwidth product exists on a source without proper documentation, so we are not sure if it qualifies as an "objective copy" or not. This quite often happens when a record manufacturer has had privileged access to metal-parts or vinyl pressings or master-tapes. He may have used them specifically for a modern reissue, but given the reissue subjective treatment using undocumented trade secrets, so we cannot reverse-engineer it to get an objective copy. However, even if we have a poor "original," we can use it to see whether the reissue qualifies as an "objective copy" or not. I call this the "partially objective" copy. Straightforward comparison with a changeover switch is usually sufficient to determine whether the new version is "objective." If the manufacturer hasn't added irreversible effects, we may even be able to re-equalise or alter the speed of his version to match the original, and achieve a better end-result. To assess the re-equalisation objectively, we may need to compare the two versions with a spectrum analyser or use a Thorn-EMI "Aquaid" System (Ref. 2). All this underlines the need for rigorous discographical research before the session.

The Power-Bandwidth Principle shows quite unambiguously the advantages of *not* copying a recording if we don't have to. Furthermore, an analogue original will always contain a certain amount of extra information hidden in it which may become available to future technologists. It will often be lost if we copy the original, even if we use the best technology we have. The late talented engineer Michael Gerzon (Ref. 3) claimed that over 99% of the information may be thrown away. Personally I consider this an overestimate; but I could agree to its being in the order of 25%. The difference may partly be because we have different definitions of the word "information." But, either way, Gerzon's message agrees with mine - KEEP THE ORIGINALS.

A final point is that it goes without saying that the facilities for cleaning originals, and otherwise restoring them to a ready-to-play state, must be provided (see Appendix 1).

2.9 Documentation strategy

I will not dwell upon the well-known empirical rule, confirmed upon numerous occasions, that it takes at least twice as long to document a recording as it does to play it. Note that I'm only talking about documenting the recorded contents now, not the technical features!

Personally, I am rather attracted by the idea that there should be a version of the documentation on the copy itself. This means that as long as the recording survives, so does the documentation, and the two can never be separated. In olden times this was achieved by a spoken announcement, and there is much to be said for this technique; as long as the copy is playable, it can be identified. For really long term purposes I consider such an announcement should be made by an expert speaker, since questions of pronunciation will arise in future years.

On the other hand, a spoken announcement isn't "machine-readable." With a digital copy the documentation might be stored as ASCII text, or as a facsimile of a written document; but as yet I have no practical experience of these techniques. There are

(unfortunately) several "standard" proposals for storing such "metadata" in digital form. And the end of Section 3.7 will warn you of potential problems with this idea.

As we proceed through this manual, we shall see that technical documentation could also become very complex, and the strategy for your archive will largely depend on what has gone before. My former employer, the BBC, always had a paper "recording report" accompanying every radio recording. It was useless without one, because it had to have the producer's signature to say it was ready for transmission before the network engineer would transmit it. But my current employer, the British Library Sound Archive, does not have such a system. It's probably too late to introduce it, because the idea of a recording-report can only work if alarm-bells ring in its absence.

By adding technical documentation, I don't wish it to take even longer to document a recording. This is especially important, because a technical report can only be completed by technical staff, and if both the operators and the equipment are unproductive while this goes on, it is very expensive. My suggested alternative is to establish a standard set of procedures, called something simple like "X1", and simply write down "Copied to procedure X1" followed by the operator's signature. (I consider the signature is important, since *only* the operator can certify that the copy is a faithful representation of the original).

This implies that "Procedure X1" is documented somewhere else, and here we must face the possibility that it may become lost. This actually happened to both my employers. When I was in the BBC the change from CCIR to IEC tape-recording equalisation at 19cm/sec took place (section 7.8), and was implemented immediately on quarter-inch tape; but not immediately on 16mm sepmag film, which happens to run at the same speed. For the next year I got complaints that my films sounded "woolly" on transmission, despite rigorous calibration of the equipment and my mixing the sound with more and more treble. When the truth dawned I was very angry; the problem (and damage to my reputation) could have been avoided by proper documentation. I was determined this should not happen when I joined the Sound Archive. Unhappily, a new director was once appointed who decided there was too much paperwork about, and scrapped most of it. The result is that, to this day, we do not know how-and-when the change to IEC equalisation took place at the Archive, so we often cannot do objective copies.

The two problems of "technical rehearsal" and "time to do the documentation" might both be solved by a system of rehearsals before the transfers actually take place. Thus, a working day might consist of the operator and the discographer working together in a low-tech area to decide on such things as playing-speeds, the best copy or copies to transfer, and whether alternatives are the same or not. The catalogue-entry can be started at the same time, and perhaps spoken announcements can be pre-recorded. There is also time to research anything which proves to need investigation. The actual "high-tech" transfers could then take place at a later date with much greater efficiency.

The advantages and disadvantages of converting analogue sounds to digital are the subject of the next chapter. We shall learn that there are some processes which should always be carried out in the analogue domain - speed-setting, for example - and some best carried out in the digital domain - various types of noise reduction, for example. Thus the overall strategy must take the two technologies into account, so that the appropriate processes happen in the right order. Also digital recordings are often inconvenient for "service copies." It may be necessary to put service-copies onto analogue media to make it easier to find excerpts, or because analogue machinery is more familiar to users.

This writer happens to believe that the analogue and digital processes should be carried out by the same people as far as possible. Up till now, digital signal processing has been rather expensive, and has tended to be farmed out to bureau services as resources permit. Not only are there communications problems and unpredictable delays which inhibit quality-checking, but a vital feedback loop - of trying something and seeing how it sounds - is broken. The overall strategy should keep the analogue and digital processes as close together as possible, although the special skills of individuals on one side of the fence or the other should not be diluted.

2.10 Absolute phase

The next few chapters of this manual will outline the different techniques for copying sounds, so I shall not deal with them here. But there are three considerations which affect *all* the techniques, and this is the only logical place to discuss them.

At various times in history, there have been debates whether a phenomenon called "absolute phase" is significant. Natural sounds consist of alternating sound pressures and rarefactions. It is argued that positive pressures should be provided at the listener's ear when positive pressures occurred at the original location, and not replaced with rarefactions. Many experienced listeners claim that when this is done correctly, the recording sounds much more satisfactory than when the phases are reversed; others claim they cannot hear any difference. I freely admit I am in the latter category; but I can see that the advantages of "absolute phase" could well exist for some people, so I should advise the sound archivist to bear this in mind and ensure all his equipment is fitted with irreversible connectors and tested to ensure absolute phase is preserved.

Since the earliest days of electrical recording, the effect has been so subtle that most equipment has been connected in essentially random ways. Furthermore, bi-directional microphones cannot have "absolute phase," because the absolute phases of artists on the two opposite sides of the microphone are inherently dissimilar. But this doesn't apply to *acoustic* recordings. As sound-pressures travelled down the horn, they resulted in the groove deviating from its path *towards the edge* of a lateral-cut disc record and *going less deep* on a hill-and-dale master-recording (due to the lever mechanisms between diaphragms and cutters). Thus, we actually *know* the absolute phase of most acoustic recordings - those which have never been copied, anyway. I therefore suggest that the copying strategy for *all* discs and cylinders should follow the convention that movements towards the disc edge for a lateral stylus and upwards movements for a hill-and-dale stylus should result in positive *increases* in the value of the digital bits. This won't mean that absolute phase is preserved in all electrical recordings, because electrical recording components were connected in random ways; but it will ensure that this aspect of the originals is preserved on the copies, whether it ever proves to be critical or not.

The English Decca Record Company has recently adopted the standard that positive-going pressures at the microphone should be represented by positive-going digits in a digital recording, and this idea is currently under discussion for an AES Standard. It seems so sensible that I advise everyone else to adopt the same procedure when planning a new installation. There is also a convention for analogue tape (Ref. 4). But virtually no-one has used it, and there is no engineering reason why the absolute phase of a tape-recording should be preserved on a copy when it is only a subjective judgement. Yet because there is a standard, archives should follow it.

2.11 Relative phase

It is even possible to enlarge upon the above ideal, and insist on correct "relative phase" as well. Please allow me to explain this, even though we shan't encounter the problem very often. Practical recording equipment (both analogue and digital) introduces relative phase shifts between different components of the same sound, which may occur due to acoustic effects, mechanical effects, or electronic effects. Any piece of equipment which "rolls off" the extreme high frequencies, for example, also delays them with respect to the low frequencies - admittedly by not very much, half a cycle at most. Since this happens every time we shout through a wall (for example), our ears have evolved to ignore this type of delay.

Many years ago at my Engineering Training School, our class was given a demonstration which was supposed to prove we couldn't hear relative phase distortion. The test-generator comprised eight gearwheels on a common axle. The first had 100 teeth, the second 200, the third 300, etc. As the axle rotated, eight pickup coils detected each tooth as it passed. The eight outputs were mixed together, displayed on an oscilloscope, and reproduced on a loudspeaker. The pickup coils could be moved slightly in relation to the gearwheels. As this was done, the relative phases of the components changed, and the waveform displayed on the oscilloscope changed radically. The sound heard from the loudspeaker wasn't supposed to change; but of course there was one sceptic in our class who insisted it did, and when the class had officially finished, we spent some time in the lab blind-testing him - the result was that he could indeed hear a difference.

But I don't mention this for the small proportion of listeners who can hear a difference. I mention it because the elimination of overload distortion may depend critically upon the correct reproduction of "relative phase." So I shall be insisting on reproduction techniques which have this feature, and on using originals (since we usually don't know the relative-phase characteristics of any equipment making copies).

2.12 Scale distortion

The third consideration has had several names over the years. The controversy flared up most brightly in the early 1960s, when it was called "scale distortion". It arises from the fact that we almost never hear a sound recording at the same volume as the original sounds. Various psychoacoustic factors come into this, which I won't expound now, but which may be imagined by knowledgeable readers when I mention the "Fletcher-Munson Curves." Where does the controversy arise? Because it is not clear what we should do when the sounds *are* reproduced at the wrong volume. I think everyone agrees that in the ideal world we *should* reproduce the original volume. The trouble for archivists is that we do not usually have objective knowledge of what this original volume was. A standard sound-calibration would be needed at every location, and this would have to be included on every recording. Such calibrations do occasionally appear on recordings of industrial noises or historic musical instruments, but they are the exception rather than the rule. Yet every time we have even a tiny scrap of such information we should preserve it.

The acoustic-recording system is again a case where this applies. It was not possible to alter the sensitivity of an acoustic recording machine during a "take", so it would be silly to transfer such a recording without including a calibration signal to link the original waveform with the transferred version. And I would point out that many early

commercial electrically-recorded discs were subject to "wear tests" before being approved for publication. At least one studio kept documentary evidence of the settings of their equipment, in case they were forced to retake an item because a test-pressing wore out (Section 6.19). Thus we do have an indication of how loud the performance was, although we have not yet learnt how to interpret this information.

2.13 Conclusion

Unfortunately, in the real world, procedures cannot be perfect, and ad-hoc decisions frequently have to be made. In the remainder of this manual, we shall see there are many areas for which technical information is incomplete. We must avoid making any objective copies unless we have the technology and the knowledge to do them. There must also be a deliberate and carefully-constructed policy of what to do in less-than-ideal circumstances. For example, in today's world with very restricted facilities, which should have maximum priority: vulnerable media, media in maximum demand, or media which could result in further knowledge? What should be the policy if the archive *cannot* get hold of a good copy of a record? What should be the policy if appropriate technology is not available? And is this decision affected when fundamental engineering theory tells us such technology is *always* impossible?

Fortunately, there's more than enough work for my recommendations to be implemented immediately. We do not have to wait for answers to those questions!

REFERENCES

1: Mark Buchanan, "Beyond reality" (*article*), London: *New Scientist* Vol. 157 No. 2125 (14th March 1998), pp. 27-30. A more general article is: Robert Matthews, "I Is The Law" (*article*), London: *New Scientist* Vol. 161 No. 2171 (30th January 1999), pp. 24-28.

2: Richard Clemow, "Computerised tape testing" (*article*), London: *One to One* (magazine) No. 53 (September 1994), pp. 67-75.

3: Michael Gerzon, "Don't Destroy The Archives!". A technical report, hitherto unpublished, dated 14th December 1992.

4: Lipshitz and Vanderkooy, "Polarity Calibration Tape (Issue 2)" (*article*), Journal of the Audio Engineering Society Vol. 29 Nos. 7/8 (July/August 1981), pp. 528-530.

3 Digital conversion of analogue sound

3.1 The advantages of digital audio

There is a primary reason why digital recordings appeal to sound archivists. Once digital encoding has been achieved, they can in principle last for ever without degradation, because digital recordings can in principle be copied indefinitely without suffering any further loss of quality. This assumes: (1) the media are always copied in the digital domain before the errors accumulate too far, (2) the errors (which can be measured) always lie within the limits of the error-correction system (which can also be measured), and (3) after error-correction has been achieved, the basic digits representing the sound are not altered in any way. When a digital recording goes through such a process successfully, it is said to be "cloned."

Both in theory and in practice, no power-bandwidth product (Section 2.3) is lost when cloning takes place - there can be no further loss in quality. However, it also means the initial analogue-to-digital conversion must be done well, otherwise faults will be propagated forever. In fact, the Compact Digital Disc (CD) has two "layers" of error-correction, and (according to audio folk-culture) the format was designed to be rugged enough to allow a hole one-sixteenth of an inch diameter (1.5mm) to be drilled through the disc without audible side-effects.

For all these reasons, the word "digital" began to achieve mystical qualities to the general public, many of whom evidently believe that anything "digital" must be superior! I am afraid much of this chapter will refute that idea. It will also be necessary to understand what happens when a recording is copied "digitally" without actually being "cloned".

The power-bandwidth products of most of today's linear pulse-code modulation media exceed most of today's analogue media, so it seems logical to copy all analogue recordings onto digital carriers anyway, even if the digital coding is slightly imperfect. But we must understand the weaknesses of today's systems if we are to avoid them (thus craftsmanship is still involved!), and we should ideally provide test-signals to document the conversion for future generations.

If you are a digital engineer, you will say that digital pulse-code modulation is a form of "lossless compression", because we don't have to double the ratio between the powers of the background-noise and of the overload-point in order to double the power-bandwidth product. In principle, we could just add one extra digital bit to each sample, as we shall see in the next section. Now I am getting ahead of myself; but I mention this because there is sometimes total lack of understanding between digital and analogue engineers about such fundamental issues. I shall therefore start by describing these fundamental issues very thoroughly. So I must apologise to readers on one side of the fence or the other, for apparently stating the obvious (or the incomprehensible).

A digital recording format seems pretty idiot-proof; the data normally consists of ones and zeros, with no room for ambiguity. But this simply isn't the case. *All* digital carriers store the digits as *analogue* information. The data may be represented by the size of a pit, or the strength of a magnetic domain, or a blob of dye. All these are quantified using *analogue* measurements, and error-correction is specifically intended to get around this difficulty (so you don't have to be measuring the size of a pit or the strength of a tiny magnet).

Unfortunately, such misunderstandings even bedevil the process of choosing an adequate medium for storing the digitised sound. There are at least two areas where these misunderstandings happen. First we must ask, are tests based on *analogue* or *digital* measurements (such as "carrier-to-noise ratio" or "bit error-rates")? And secondly, has the digital reproducer been *optimised* for reproducing the analogue features, or is it a self-adjusting "universal" machine (and if so, how do we judge *that*)?

Finally, the word "format" even has two meanings, both of which should be specified. The digits have their own "format" (number of bits per sample, sampling-frequency, and other parameters we shall meet later in this chapter); and the actual digits may be recorded on either analogue or digital carrier formats (such as Umatic videocassettes, versions of Compact Digital discs, etc.). The result tends to be total confusion when "digital" and "analogue" operators try communicating!

3.2 Technical restrictions of digital audio - the "power" element

I shall now look at the principles of digital sound recording with the eyes of an analogue engineer, to check how digital recording can conserve the power-bandwidth product. I will just remind you that the "power" dimension defines the interval between the faintest sound which can be recorded, and the onset of overloading.

All digital systems overload abruptly, unless they are protected by preceding analogue circuitry. Fortunately analogue sound recordings generally have a smaller "power" component to their power-bandwidth products, so there is no need to overload a digital medium when you play an analogue recording. Today the principal exception concerns desktop computers, whose analogue-to-digital converters are put into an "electrically noisy" environment. Fortunately, very low-tech analogue noise-tests define this situation if it occurs; and to make the system "idiot-proof" for non-audio experts, consumer-quality cards often contains an automatic volume control as well (chapter 10).

Unfortunately, we now have two philosophies for archivists, and you may need to work out a policy on this matter. One is to transfer the analogue recording at "constant-gain," so future users get a representation of the signal-strength of the original, which might conceivably help future methods of curing analogue distortion (section 4.15). The other is that all encoders working on the system called Pulse Code Modulation (or PCM) have defects at the low end of the dynamic range. Since we are copying (as opposed to doing "live" recording), we can predict very precisely what the signal volume will be before we digitise it, set it to reach the maximum undistorted volume, and drown the quiet side-effects.

These low-level effects are occupying the attention of many talented engineers, with inevitable hocus-pocus from pundits. I shall spend the next few paragraphs outlining the problem. If an ideal solution ever emerges, you will be able to sort the wheat from the chaff and adopt it yourself. Meanwhile, it seems to me that any future methods of reducing overload distortion will have to "learn" - in other words, adapt themselves to the actual distortions present - rather than using predetermined "recipes".

All PCM recordings will give a "granular" sound quality if they are not "dithered." This is because wanted signals whose volume is similar to the least significant bit will be "chopped up" by the lack of resolution at this level. The result is called "quantisation distortion." One solution is always to have some background noise.

Hiss may be provided from a special analogue hiss-generator preceding the analogue-to-digital converter. (Usually this is there whether you want it or not!)

Alternatively it may be generated by a random-number algorithm in the digital domain. Such "dither" will completely eliminate this distortion, at the cost of a very faint steady hiss being added. The current debate is about methods of reducing, or perhaps making deliberate use of, this additional hiss.

Today the normal practice is to add "triangular probability distribution" noise, which is preferable to the "rectangular probability distribution" of earlier days, because you don't hear the hiss vary with signal volume (an effect called "modulation noise"). Regrettably, many "sound cards" for computers still use rectangular probability distribution noise, illustrating the gulf which can exist between digital engineers and analogue ones! It also illustrates why *you* must be aware of basic principles on *both* sides of the fence.

Even with triangular probability distribution noise, in the past few years this strategy has been re-examined. There are now several processes which claim to make the hiss less audible to human listeners while simultaneously avoiding quantisation distortion and modulation noise. These processes have different starting-points. For example, some are for studio recordings with very low background noise (at the "twenty-bit level") so they will sound better when configured for sixteen-bit Compact Discs. Such hiss might subjectively be fifteen decibels quieter, yet have an un-natural quality; so other processes aim for a more "benign" hiss. A process suggested by Philips adds something which sounds like hiss, but which actually comprises pseudo-random digits which carry useful information. Another process (called "auto-dither") adds pseudo-random digits which can be subtracted upon replay, thereby making such dither totally inaudible, although it will still be reproduced on an ordinary machine. Personally I advocate good old triangular probability distribution noise for the somewhat esoteric reason that it's always possible to say where the "original sound" stopped and the "destination medium" starts.

All this is largely irrelevant to archivists transferring analogue recordings to digital, except that you should not forget "non-human" applications such as wildlife recording. There is also a risk of unsuspected cumulative build-ups over several generations of digital processing, and unexpected side-effects (or actual loss of information) if some types of processing are carried out upon the results.

If you put a recording through a digital process which drops the volume of some or all of the recording so that the remaining hiss (whether "dither" or "natural") is less than the least-significant bit, it will have to be "re-dithered." We should also perform re-dithering if the resulting sounds involve fractions of a bit, rather than integers; I am told that a gain change of a fraction of a decibel causes endless side-effects! One version of a widely-used process called The Fast Fourier Transform splits the frequency-range into 2048 slices, so the noise energy may be reduced by a factor of 2048 (or more) in some slices. If this falls below the least-significant bit, quantisation distortion will begin to affect the wanted signal.

In my personal view, the best way of avoiding these troubles is to use 24-bit samples during the generation of any archive and objective copies which need digital signal processing. The results can then be reduced to 16-bits for storage, and the side-effects tailored at that stage. Practical 24-bit encoders do not yet exist, because they have insufficiently low background noise; but this is exactly why they solve the difficulty. Provided digital processing takes place in the 24-bit domain, the side-effects will be at least 48 decibels lower than with 16-bit encoding, and quantisation distortion will hardly ever come into the picture at all.

On the other hand, if 16-bit processing is used, the operator must move his sound from one process to the next *without* re-dithering it unnecessarily (to avoid building up

the noise), but to *add* the dither whenever it is needed to kill the distortion. This means intelligent judgement throughout. The operator must ask himself "Did the last stage result in part or all of the wanted signal falling below the least-significant bit?" And at the end of the processes he must ask himself "Has the final stage resulted in part or all of the signal falling below the least-significant bit?" If the answer to either of these questions is Yes, then the operator must carry out a new re-dithering stage. This is an argument for ensuring the same operator sees the job through from beginning to end.

3.3 Technical limitations of digital audio: the "bandwidth" element

In this section I shall discuss how the frequency range may be corrupted by digital encoding. The first point is that the coding system known as "PCM" (almost universally used today) requires "anti-aliasing filters." These are deliberately introduced to reduce the frequency range, contrary to the ideals mentioned in section 2.4.

This is because of a mathematical theorem known as "Shannon's Sampling Theorem." Shannon showed that if you have to take samples of a time-varying signal of any type, the data you have to store will correctly represent the signal if the signal is frequency-restricted *before* you take the samples. After this, you only need to sample the amplitude of the data at twice the cut-off frequency. This applies whether you are taking readings of the water-level in a river once an hour, or encoding high-definition television pictures which contain components up to thirty million per second.

To describe this concept in words, if the "float" in the river has "parasitic oscillations" due to the waves, and bobs up and down at 1Hz, you will need to make measurements of its level at least twice a second to reproduce all its movements faithfully without errors. If you try to sample at (say) only once a minute, the bobbing-actions will cause "noise" added to the wanted signal (the longer-term level of the river), reducing the precision of the measurements (by reducing the power-bandwidth product).

Any method of sampling an analogue signal will misbehave if it contains any frequencies higher than half the sampling frequency. If, for instance, the sampling-frequency is 44.1kHz (the standard for audio Compact Discs), and an analogue signal with a frequency of 22.06kHz gets presented to the analogue-to-digital converter, the resulting digits will contain this frequency "folded back" - in this case, a spurious frequency of 22.04kHz. Such spurious sounds can never be distinguished from the wanted signal afterwards. This is called "aliasing."

Unfortunately, aliasing may also occur when you conduct "non-linear" digital signal-processing upon the results. This has implications for the processes you should use before you put the signal through the analogue-to-digital conversion stage. On the other hand, quite a few processes are easier to carry out in the digital domain. These processes must be designed so as not to introduce significant aliasing, otherwise supposedly-superior methods may come under an unquantifiable cloud - although most can be shown up by simple low-tech analogue tests!

The problem is Shannon's sampling theorem again. For example, a digital process may recognise an analogue "click" because it has a leading edge and a trailing edge, both of which are faster than any naturally-occurring transient sound. But Shannon's sampling theorem says the *frequency* range must not exceed half the sampling-frequency; this bears no simple relationship to the *slew-rate*, which is what the computer will recognise in this example. Therefore the elimination of the click will result in aliased artefacts mixed up

with the wanted sound, unless some very computation-intensive processes are used to bandwidth-limit these artefacts.

Because high-fidelity digital audio was first tried in the mid-1970s when it was very difficult to store the samples fast enough, the anti-aliasing filters were designed to work just above the upper limit of human hearing. There simply wasn't any spare capacity to provide any "elbow-room," unlike measuring the water-level in a river. The perfect filters required by Shannon's theorem do not exist, and in practice you can often hear the result on semi-pro or amateur digital machines if you try recording test-tones around 20-25 kHz. Sometimes the analogue filters will behave differently on the two channels, so stereo images will be affected. And even if "perfect filters" are approached by careful engineering, another mathematical theorem called "the Gibbs effect" may distort the resulting waveshapes.

An analogue "square-wave" signal will acquire "ripples" along its top and bottom edges, looking exactly like a high-frequency resonance. If you are an analogue engineer you will criticise this effect, because analogue engineers are trained to eliminate resonances in their microphones, their loudspeakers, and their electrical circuitry; but this phenomenon is actually an artefact of the mathematics of steeply bandwidth-limiting a frequency before you digitise it. Such factors cause distress to golden-eared analogue engineers, and have generated much argument against digital recording. "Professional-standard" de-clicking devices employ "oversampling" to overcome the Gibbs effect on clicks; but this cannot work with declicking software on personal computers, for example.

The Gibbs effect can be reduced by reasonably gentle filters coming into effect at about 18kHz, when only material above the limit of hearing for an adult human listener would be thrown away. But we might be throwing away information of relevance in other applications, for instance analysis by electronic measuring instruments, or playback to wildlife, or to young children (whose hearing can sometimes reach 25kHz). So we must first be sure only to use linear PCM at 44.1kHz when the subject matter is *only* for adult human listeners. This isn't a condemnation of digital recording as such, of course. It is only a reminder to use the correct tool for any job.

You can see why the cult hi-fi fraternity sometimes avoids digital recordings like the plague! Fortunately, this need not apply to you. Ordinary listeners cannot compare quality "before" and "after"; but you can (and should), so you needn't be involved in the debate at all. If there is a likelihood of mechanical or non-human applications, then a different medium might be preferable; otherwise you should ensure that archive copies are made by state-of-the-art converters checked in the laboratory and double-checked by ear.

I could spend some time discussing the promise of other proposed digital encoding systems, such as non-linear encoding or delta-sigma modulation, which have advantages and disadvantages; but I shall not do so until such technology becomes readily available to the archivist. One version of delta-sigma modulation has in fact just become available; Sony/Philips are using it for their "Super Audio CD" (SACD). The idea is to have "one-bit" samples taken at very many times the highest wanted frequency. Such "one-bit samples" record whether the signal is going "up" or "down" at the time the sample is taken. There is no need for an anti-aliasing filter, because Shannon's theorem does not apply. However, the process results in large quantities of quantisation noise above the upper limit of human hearing. At present, delta-modulation combines the advantage of no anti-aliasing filter with the disadvantage that there is practically no signal-processing technology which can make use of the bitstream. If delta modulation "takes off", signal processes will eventually become available, together with the technology for storing the

required extra bits. (SACD needs slightly more bits than a PCM 24-bit recording sampled at 96kHz). But for the moment, we are stuck with PCM, and I shall assume PCM for the remainder of this manual.

3.4 Operational techniques for digital encoding

I hesitate to make the next point again, but I do so knowing many operators don't work this way. Whatever the medium, whether it be digital or analogue, the transfer operator *must* be able to compare source and replay upon a changeover switch. Although this is normal for analogue tape recording, where machines with three heads are used to play back a sound as soon as it is recorded, they seem to be very rare in digital environments. Some machines do not even give "E-E monitoring", where the encoder and decoder electronics are connected back-to-back for monitoring purposes. So, I think it is absolutely *vital* for the operator to *listen* when his object is to copy the wanted sound without alteration. Only *he* is in a position to judge when the new version is faithful to the original, and he must be prepared to sign his name to witness this. Please remember my philosophy, which is that all equipment must give satisfactory measurements; but the ear must be the *final* arbiter.

Even if E-E monitoring passes this test, it does not prove the copy is perfect. There could be dropouts on tape, or track-jumps on CD-R discs. Fortunately, once digital conversion has been done, automatic devices may be used to check the medium for errors; humans are not required.

I now mention a novel difficulty, documenting the performance of the anti-aliasing filter and the subsequent analogue-to-digital converter. Theoretically, everything can be documented by a simple "impulse response" test. The impulse-response of a *digital-to-analogue* converter is easy to measure, because all you need is a single sample with all its bits at "1". This is easy to generate, and many test CDs carry such signals. But there isn't an "international standard" for documenting the performance of analogue-to-digital converters. One is desperately needed, because such converters may have frequency-responses down to a fraction of 1Hz, which means that one impulse may have to be separated by many seconds; while the impulse must also be chosen with a combination of duration and amplitude to suit the "sample-and-hold" circuit as well as not to overload the anti-aliasing filter.

At the British Library Sound Archive, we use a Thurlby Thandar Instruments type TGP110 analogue Pulse Generator in its basic form (not "calibrated" to give highly precise waveforms). We have standardised on impulses exactly 1 microsecond long. The resulting digitised shape can be displayed on a digital audio editor.

3.5 Difficulties of "cloning" digital recordings

For the next three sections, you will think I am biased against "digital sound"; but the fact that many digital formats have higher power-bandwidth products should not mean we should be unaware of their problems. I shall now point out the defects of the assumption that digital recordings can be cloned. My earlier assertion that linear PCM digital recordings can be copied without degradation had three hidden assumptions, which seem increasingly unlikely with the passage of time. They are that the sampling frequency, the pre-emphasis, and the bit-resolution remain constant.

If (say) it is desired to copy a digital PCM audio recording to a new PCM format with a higher sampling frequency, then either the sound must be converted from digital to analogue and back again, or it must be recalculated digitally. When the new sampling-frequency and the old have an arithmetical greatest common factor *n*, then every *n*th sample remains the same; but all the others must be a weighted average of the samples either side, and this implies that the new bitstream must include "decimals." For truthful representation, it cannot be a stream of integers; rounding-errors (and therefore quantisation-distortion) are bound to occur.

The subjective effect is difficult to describe, and with practical present-day systems it is usually inaudible when done just once (even with integers). But now we have a quite different danger, because once people realise digital conversion is possible (however imperfect), they will ask for it again and again, and errors will accumulate through the generations unless there is strict control by documentation. Therefore it is necessary to outguess posterity, and choose a sampling-frequency which will not become obsolete.

For the older audio media, the writer's present view is that 44.1kHz will last, because of the large number of compact disc players, and new audio media (like the DCC and MiniDisc) use 44.1kHz as well. I also consider that for the media which need urgent conservation copying (wax cylinders, acetate-based recording tape, and "acetate" discs, none of which have very intense high frequencies), this system results in imperceptible losses, and the gains outweigh these. For picture media I advocate 48kHz, because that is the rate used by digital video formats.

But there will inevitably be losses when digital sound is moved from one domain to the other, and this will get worse if sampling-frequencies proliferate. Equipment is becoming available which works at 96kHz, precisely double the frequency used by television. Recordings made at 48kHz can then be copied to make the "even-numbered" samples, while the "odd-numbered" samples become the averages of the samples either side. The options for better anti-aliasing filters, applications such as wildlife recording, preservation of transients (such as analogue disc clicks), transparent digital signal-processing, etc. remain open. Yet even this option requires that we document what has happened - future workers cannot be expected to guess it.

Converting to a *lower* sampling frequency means that the recording must be subjected to a new anti-aliasing filter. Although this filtering can be done in the digital domain to reduce the effects of practical analogue filters and two converters, it means throwing away some of the information of course.

The next problem is pre-emphasis. This means amplifying some of the wanted high frequencies before they are encoded, and doing the inverse on playback. This renders the sound less liable to quantisation distortion, because any "natural" hiss is about twelve decibels stronger. At present there is only one standard pre-emphasis characteristic for digital audio recording (section 7.3), so it can only be either ON or OFF. A flag is set in the digital data-stream of standardised interconnections, so the presence of pre-emphasis may be recognised on playback. And there is a much more powerful pre-emphasis system (the "C.C.I.T" curve) used in telecommunications. It is optimised for 8-bit audio work, and 8 bits were once often used by personal computers for economical sound recording. Hopefully my readers won't be called upon to work with CCIT pre-emphasis, because digital sound-card designers apparently haven't learnt that 8-bit encoders could then give the dynamic range of professional analogue tape; but you ought to know the possibility exists! But if the recording gets copied to change its pre-emphasis status, whether through a digital process or an analogue link, some of the power-bandwidth product will be lost each time. Worse still, some digital signal devices (particularly hard-disk editors)

strip off the pre-emphasis flag, and it is possible that digital recordings may be reproduced incorrectly after this. (Or worse still, *parts* of digital recordings will be reproduced incorrectly).

I advise the reader to make a definite policy on the use of pre-emphasis and stick to it. The "pros" are that with the vast majority of sounds meant for human listeners, the power-bandwidth product of the medium is used more efficiently; the "cons" are that this doesn't apply to most animal sounds, and digital metering and processing (Chapter 3) are sometimes more difficult. Yet even this option requires that we document what has happened - future workers cannot be expected to guess it.

Converting a linear PCM recording to a greater number of bits (such as going from 14-bit to 16-bit) does not theoretically mean any losses. In fact, if it happens at the same time as a sample-rate conversion, the new medium can be made to carry two bits of the decimal part of the interpolation mentioned earlier, thereby reducing the side-effects. Meanwhile the most significant bits retain their status, and the peak volume will be the same as for the original recording. So if it ever becomes normal to copy from (say) 16-bits to 20-bits in the digital domain, it will be easier to change the sampling frequency as well, because the roundoff errors will have less effect, by a factor of 16 in this example. Thus, to summarise, satisfactory sampling-frequency conversion will always be difficult; but it will become easier with higher bit-resolutions. Yet even this option requires that we document what has happened - future workers cannot be expected to guess it.

All these difficulties are inherent - they cannot be solved with "better technology" - and this underlines the fact that analogue-to-digital conversions and digital signal processing must be done to the highest possible standards. Since the AES Interface for digital audio allows for 24-bit samples, it seems sensible to plan for this number of bits, even though the best current converters can just about reach the 22-bit level.

It is nearly always better to do straight digital standards conversions in the digital domain when you must, and a device such as the Digital Audio Research "DASS" Unit may be very helpful. This offers several useful processes. Besides changing the pre-emphasis status and the bit-resolution, it can alter the copy-protect bits, reverse both relative and absolute phases, change the volume, and remove DC offsets. The unit automatically looks after the process of "re-dithering" when necessary, and offers two different ways of doing sampling-rate conversion. The first is advocated when the two rates are not very different, but accurate synchronisation is essential. This makes use of a buffer memory of 1024 samples. When this is either empty or full, it does a benign digital crossfade to "catch up," but between these moments the data-stream remains uncorrupted. The other method is used for widely-differing sampling-rates which would overwhelm the buffer. This does the operation described at the start of this section, causing slight degradation throughout the whole of the recording. Yet even this option requires that we document what has happened - future workers cannot be expected to guess it.

Finally I must remind you that digital recordings are not necessarily above criticism. I can think of many compact discs with quite conspicuous analogue errors on them. Some even have the code "DDD" (suggesting only digital processes have been used during their manufacture). It seems some companies use analogue noise reduction systems (Chapter 8) to "stretch" the performance of 16-bit recording media, and they do not understand the "old technology." Later chapters will teach you how to get accurate sound from analogue media; but please keep your ears open, and be prepared for the *same* faults on digital media!

3.6 Digital data compression

The undoubted advantages of linear PCM as a way of storing audio waveforms are being endangered by various types of digital data compression. The idea is to store digital sound at lower cost, or to transmit it using less of one of our limited natural resources (the electromagnetic spectrum).

Algorithms for digital compression are of two kinds, "lossless" and "lossy." The lossless ones give you back the same digits after decompression, so they do not affect the sound. There are several processes; one of the first (Compusonics) reduced the data to only about four-fifths the original amount, but the compression rate was fixed in the sense that the same number of bits was recorded in a particular time. If we allow the recording medium to vary its data-rate depending on the subject matter, data-reduction may be two-thirds for the worst cases to one-quarter for the best. My personal view is that these aren't worth bothering with, unless you're consistently in the situation where the durations of your recordings are fractionally longer than the capacity of your storage media.

For audio, some lossless methods actually make matters worse. Applause is notorious for being difficult to compress; if you *must* use such compression, test it on a recording of continuous applause. You may even find the size of the file *increases*.

But the real trouble comes from lossy systems, which can achieve compression factors from twofold to at least twentyfold. They all rely upon psychoacoustics to permit the digital data stream to be reduced. Two such digital sound recording formats were the Digital Compact Cassette (DCC) and the MiniDisc, each achieving about one-fifth the original number of bits; but in practice, quoted costs were certainly not one-fifth! While they make acceptable noises on studio-quality recordings, it is very suspicious that no "back-catalogue" is offered. The unpredictable nature of background noise always gives problems, and that is precisely what we find ourselves trying to encode with analogue sources. Applause can also degenerate into a noisy "mush". The real reason for their introduction was not an engineering one. Because newer digital systems were designed so they could not clone manufactured CDs, the professional recording industry was less likely to object to their potential for copyright abuse (a consideration we shall meet in section 3.8 below).

Other examples of digital audio compression methods are being used for other applications. To get digital audio between the perforation-holes of 35mm optical film, cinema surround-sound was originally coded digitally into a soundtrack with lossy compression. Initial reports suggested it sometimes strained the technology beyond its breaking-point. While ordinary stereo didn't sound too bad, the extra information for the rear-channel loudspeakers caused strange results to appear. An ethical point arises here, which is that the sound-mixers adapted their mixing technique to suit the compression-system. Therefore the sound was changed to suit the medium. (In this case, no original sound existed in the first place, so there wasn't any need to conserve it.)

A number of compression techniques are used for landline and satellite communication, and here the tradeoffs are financial - it costs money to buy the power-bandwidth product of such media. Broadcasters use digital compression a lot – NICAM stereo and DAB have it - but this is more understandable, because there is a limited amount of electromagnetic spectrum which we must all share, especially for consistent reception in cars. At least we can assume that wildlife creatures or analytical machinery won't be listening to the radio, visiting cinemas, or driving cars.

The "advantages" of lossy digital compression have six counterarguments. (1) The GDR Archive found that the cost of storage is less than five percent of the total costs of running an archive, so the savings are not great; (2) digital storage (and transmission) are set to get cheaper, not more expensive; (3) even though a system may sound transparent now, there's no evidence that we may not hear side-effects when new applications are developed; (4) once people think digital recordings can be "cloned", they will put lossy compression systems one after the other and think they are preserving the original sound, whereas cascading several lossy compression-systems magnifies all the disadvantages of each; (5) data compression systems will themselves evolve, so capital costs will be involved; (6) even digital storage media with brief shelf-lives seem set to outlast current compression-systems.

There can be no "perfect" lossy compression system for audio. In section 1.2 I described how individual human babies *learned* how to hear, and how a physiological defect might be compensated by a psychological change. Compression-systems are always tested by people with "normal" hearing (or sight in the case of video compression). This research may be inherently wrong for people with defective hearing (or sight). Under British law at least, the result might be regarded as discriminating against certain members of the public. Although there has not yet been any legal action on this front, I must point out the possibility to my readers.

With all lossy systems, unless cloning with error-correction is provided, the sound will degrade further each time it is copied. I consider a sound archive should have nothing to do with such media unless the ability to clone the stuff with error-correction is made available, and the destination-media and the hardware also continue to be available. The degradations will then stay the same and won't accumulate. (The DCC will do this, but won't allow the accompanying text, documentation, and start-idents to be transferred; and you have to buy a special cloning machine for MiniDisc, which erases the existence of edits).

Because there is no "watermark" to document what has happened en route, digital television is already giving endless problems to archivists. Since compression is vital for getting news-stories home quickly - and several versions may be cascaded depending on the bitrates available *en route* - there is no way of knowing which version is "nearest the original". So even this option requires that we document what has happened if we can - future workers cannot be expected to guess it.

Ideally, hardware should be made available to decode the compressed bit-stream with no loss of power-bandwidth product under audible or laboratory test-conditions. This isn't always possible with the present state-of-the-art; but unless it is, we can at least preserve the sound in the manner the misguided producers intended, and the advantages of uncompressed PCM recording won't come under a shadow. When neither of these strategies is possible, the archive will be forced to convert the original back to analogue whenever a listener requires, and this will mean considerable investment in equipment and perpetual maintenance costs. All this underlines my feeling that a sound archive should have nothing to do with lossy data compression.

My mention of the Compusonics system reminds me that it isn't just a matter of hardware. There is a thin dividing line between "hardware" and "software." I do not mean to libel Messrs. Compusonics by the following remark, but it is a point I must make. Software can be copyrighted, which reduces the chance of a process being usable in future.

3.7 A severe warning

I shall make this point more clearly by an actual example in another field. I started writing the text of this manual about ten years ago, and as I have continued to add to it and amend it, I have been forced to keep the same word-processing software in my computer. Unfortunately, I have had three computers during that time, and the word-processing software was licensed for use on only one machine. I bought a second copy with my second machine, but by the time I got my third machine the product was no longer available. Rather than risk prosecution by copying the software onto my new machine, I am now forced to use one of the original floppies in the floppy disk drive, talking to the text on the hard drive. This adds unnecessary operational hassles, and only works at all because the first computer and the last happen to have had the same (copyright) operating system (which was sheer luck).

For a computer-user not used to my way of thinking, the immediate question is "What's wrong with buying a more up-to-date word-processor?" My answer is threefold. (1) There is nothing wrong with my present system; (2) I can export my writings in a way which avoids having to re-type the stuff for another word-processor, whereas the other way round simply doesn't work; (3) I cannot see why I should pay someone to re-invent the wheel. (I have better things to spend my time and money on)! And once I enter these treacherous waters, I shall have to continue shelling out money for the next fifty years or more.

I must now explain that last remark, drawing from my own experience in Britain, and asking readers to apply the lessons of what I say to the legal situation wherever they work. In Britain, copyright in computer software lasts fifty years after its first publication (with new versions constantly pushing this date further into the future). Furthermore, under British law, you do not "buy" software, you "license" it. So the normal provisions of the Consumer Protection Act (that it must be "fit for the intended purpose" - i.e. it must *work*) simply do not apply. Meanwhile, different manufacturers are free to impose their ideas about what constitutes "copying" to make the software practicable. (At present, this isn't defined in British law, except to say that software may always legally be copied into RAM - memory in which the information vanishes when you switch off the power). Finally, the 1988 Copyright Act allows "moral" rights, which prevent anyone modifying anything "in a derogatory manner". This right cannot be assigned to another person or organisation, it stays with the author; presumably in extreme cases it could mean the licensee may not even modify it.

It is easy to see the difficulties that sound archivists might face in forty-nine years time, when the hardware has radically changed. (Think of the difficulties I've had in only ten years with text!) I therefore think it is essential to plan sound archival strategy so that *no* software is involved. Alternatively, the software might be "public-domain" or "home-grown," and ideally one should have access to the "source code" (written in an internationally-standardised language such as FORTRAN or "C"), which may subsequently be re-compiled for different microprocessors. I consider that even temporary processes used in sound restoration, such as audio editors or digital noise reduction systems (Chapter 3) should follow the same principles, otherwise the archivist is bound to be painted into a corner sooner or later. If hardware evolves at its present rate, copyright law may halt legal playback of many digital recording formats or implementation of many digital signal processes. Under British law, once the "software" is permanently stored in a device known as an "EPROM" chip, it becomes "hardware", and the problems of

copyright software evaporate. But this just makes matters more difficult if the EPROM should fail.

I apologise to readers for being forced to point out further "facts of life," but I have never seen the next few ideas written down anywhere. It is your duty to understand *all* the "Dangers of Digital", so I will warn you about more dangers of copyright software. The obvious one, which is that hardware manufacturers will blame the software writers if something goes wrong (and vice versa), seems almost self-evident; but it still needs to be mentioned.

A second-order danger is that publishers of software often make deliberate attempts to trap users into "brand loyalty". Thus, I can think of many word-processing programs reissued with "upgrades" (real or imagined), sometimes for use with a new "operating system". But such programs usually function with at least one step of "downward compatibility", so users are not tempted to cut their losses and switch to different software. This has been the situation since at least 1977 (with the languages FORTRAN66 and FORTRAN77); but for some reason no-one seems to have recognised the problem. I regret having to make a political point here; but as both the computer-magazine and book industries are utterly dependent upon *not* mentioning it, the point has never been raised among people who matter (archivists!).

This disease has spread to digital recording media as well, with many backup media (controlled by software) having only one level of downwards compatibility, if that. As a simple example, I shall cite the "three-and-a-half inch floppy disk", which exists in two forms, the "normal" one (under MS-DOS this can hold 720 kilobytes), and the "high-density" version (which can hold 1.44 megabytes). In the "high density" version the digits are packed closer together, requiring a high-density magnetic layer. The hardware should be able to tell which is which by means of a "feeler hole", exactly like analogue audiocassettes (Chapter 6). But, to cut costs, modern floppy disk drives lack any way of detecting the hole, so cannot read 720k disks. The problem is an *analogue* one, and we shall see precisely the same problem when we talk about analogue magnetic tape. Both greater amplification *and* matched playback-heads must coexist to read older formats properly.

Even worse, the downwards-compatibility situation has spread to the operating system (the software which makes a computer work at all). For example, "Windows NT" (which was much-touted as a 32-bit operating system, although no engineer would see any advantage in that) can handle 16-bit applications, but not 8-bit ones. A large organisation has this pistol held to its head with greater pressure, since if the operating system must be changed, every computer must also be changed - overnight - or data cannot be exchanged on digital media (or if they can, with meaningless error-messages or warnings of viruses). All this arises because pieces of digital jigsaw do not fit together.

To a sound archivist, the second-order strategy is only acceptable so long as the sound recordings do not change their format. If you think that an updated operating-system is certain to be better for digital storage, then I must remind you that you will be storing successive layers of problems for future generations. Use only a format which is internationally-standardised and widely-used (such as "Red Book" compact disc), and do not allow yourself to be seduced by potential "upgrades."

A "third-order" problem is the well-known problem of "vapourware." This is where the software company deliberately courts brand-loyalty by telling its users an upgrade is imminent. This has four unfavourable features which don't apply to "hardware." First, no particular delivery-time is promised - it may be two or three years away; second, the new version will need to be re-tested by users; third, operators will

have to re-learn how to use it; and almost inevitably people will then use the new version more intensely, pushing at the barriers until it too "falls over." (These won't be the responsibility of the software company, of course; and usually extra cash is involved as well). Even if the software is buried in an EPROM chip (as opposed to a removable medium which can be changed easily), this means that sound archivists must document the "version number" for any archive copies, while the original analogue recording must be preserved indefinitely in case a better version becomes available.

And there are even "fourth-order" problems. The handbook often gets separated from the software, so you often cannot do anything practical even when the software survives. Also, many software packages are (deliberately or accidentally) badly-written, so you find yourself "trapped in a loop" or something similar, and must ring a so-called "help-line" at a massive cost to your telephone bill. Even this wouldn't matter if only the software could be guaranteed for fifty years into the future . . . .

Without wishing to decry the efforts of legitimate copyright owners, I must remind you that many forms of hardware come with associated software. For example, every digital "sound card" I know has copyright software to make it work. So I must advise readers to have nothing to do with sound cards, unless they are used solely as an intermediate stage in the generation of internationally-standardised digital copies played by means of hardware alone.

3.8 Digital watermarking and copy protection.

As I write this, digital audio is also becoming corrupted by "watermarking". The idea is to alter a sound recording so that its source may be identified, whether broadcast, sent over the Internet, or whatever. Such treatment must be rugged enough to survive various analogue or digital distortions. Many manufacturers have developed inaudible techniques for adding a watermark, although they all corrupt "the original sound" in the process. As I write this, it looks as though a system called "MusiCode" (from Aris Technologies) will become dominant (Ref. 1). This one changes successive peaks in music to carry an extra encoded message. As the decoding software will have the same problems as I described in the previous section, it won't be a complete answer to the archivist's prayer for a recording containing its own identification; but for once I can see some sort of advantage accompanying this "distortion." On the other hand, it means the archivist will have to purchase (sorry, "license") a copy of the appropriate software to make any practical use of this information. And of course, professional sound archivists will be forced to license both it and *all* the other watermarking systems, in order to identify *unmodified* versions of the same sound.

Wealthy commercial record-companies will use the code to identify their products. But such identification will not identify the copyright *owner*, for example when a product is licensed for overseas sales, or the artists own the copyright in their own music. This point is developed on another page of the same Reference (Ref. 2), where a rival watermarking system is accompanied by an infrastructure of monitoring-stations listening to the airwaves for recordings with their watermarks, and automatically sending reports to a centralised agency which will notify the copyright *owners*.

I am writing this paragraph in 1999: I confidently predict that numerous "watermarking" systems will be invented, they will allegedly be tested by professional audio listeners using the rigorous A-B comparison methods I described in section 3.4, and then discarded.

All this is quite apart from a machine-readable identification to replace a spoken announcement (on the lines I mentioned in section 2.9). Yet even here, there are currently five "standards" fighting it out in the marketplace, and as far as I can see these all depend upon Roman characters for the "metadata". I will say no more.

Another difficulty is "copy protection." In a very belated attempt to restrict the digital cloning of commercial products, digital media are beginning to carry extra "copy protection bits." The 1981 "Red Book" standard for compact discs allowed these from Day One; but sales people considered it a technicality not worthy of their attention! But people in the real world - namely, music composers and performers - soon realised there was an enormous potential for piracy; and now we have a great deal of shutting of stable doors.

The Compact Disc itself carries any "copy protect" flag, and many countries now pay royalties on retail sales of recordable CDs specifically to compensate music and record publishers. (Such discs may be marketed with a distinctly misleading phrase like "For music", which ought to read "For in-copyright published music, and copyright records, only." Then, when it doesn't record anything else, there could be action under the Trades Descriptions Act.) So a professional archive may be obliged to pay the "music" royalty (or purchase "professional" CD-Rs) to ensure the copy-protect flags are not raised. Meanwhile, blank CD-ROM discs for computers (which normally permit a third layer of error-correction) can also be used for audio, when the third layer is ignored by CD players (so the result becomes "Red Book" standard with two layers). The copy-protect bit is not (at the time of writing) raised; so now another company has entered the field to corrupt this third layer of error-correction and prevent copying from on a CD-ROM drive. (Ref. 3)

Most other media (including digital audio tape and MiniDisc) add a "copy-protect" bit when recorded on an "amateur" machine, without the idea of copyright being explained at any point - so the snag only becomes apparent when *you* are asked to clone the result. The "amateur" digital connection (SP-DIF) carries start-flags as well as copy-protect flags, while the "professional" digital connection (AES3) carries neither. So if you need to clone digital recordings including their track-flags, it may be virtually impossible with many combinations of media.

A cure is easy to see - it just means purchasing the correct equipment. Add a digital delay-line with a capacity of ten seconds or so using AES connectors. Use a source-machine which does a digital count-*down* of the time before each track ends. Then the cloning can run as a background task to some other job, and meanwhile the operator can press the track-increment button manually with two types of advance-warning - visual and aural.

3.9 The use of general-purpose computers

Desktop computers are getting cheaper and more powerful all the time, so digital sound restoration techniques will become more accessible to the archivist. So far these processes are utterly dependent upon classical linear PCM coding; this is another argument against compressed digital recordings.

Unfortunately, desktop computers are not always welcomed by the audio fraternity, because most of them have whirring disk-drives and continuous cooling fans. Analogue sound-restoration frequently means listening to the faintest parts of recordings, and just one desktop computer in the same room can completely scupper this process. Finally, analogue operators curse and swear at the inconvenient controls and non-intuitive

software. The senses of touch and instant responses are very important to an analogue operator.

It is vital to plan a way around the noise difficulty. Kits are available which allow the "system box" to be out of the room, while the keyboard screen and mouse remain on the desktop. Alternatively, we might do trial sections on noise-excluding headphones, and leave the computer to crunch through long recordings during the night-time. In section 1.4 I expressed the view that the restoration operator should not be twiddling knobs subjectively. A computer running "out of real-time" forces the operator to plan his processing logically, and actually prevents subjective intervention.

This brings us to the fact that desktop computers are only just beginning to cope with real-time digital signal processing (DSP), although this sometimes implies dedicated accelerator-boards or special "bus architectures" (both of which imply special software). On the other hand, the desktop PC is an ideal tool for solving a rare technical problem. Sound archivists do not have much cash, and there aren't enough to provide a user-base for the designers of special hardware or the writers of special software. But once we can get a digitised recording into a PC and out again, it is relatively cheap to develop a tailor-made solution to a rare problem, which may be needed only once or twice in a decade.

Even so, archivists often have specialist requirements which are needed more often than that. This writer considers an acceptable compromise is to purchase a special board which will write digital audio into a DOS file on the hard disk of a PC (I am told Turtle Beach Electronics makes such a board, although it is only 16-bit capable, and requires its own software). Then special software can be loaded from floppy disk to perform special signal processing.

3.10 Processes better handled in the analogue domain

The present state-of-the-art means that all digital recordings will be subject to difficulties if we want to alter their speeds. To be pedantic, the difficulties occur when we want to change the speed of a digital *recording*, rather than its playback into the analogue domain. In principle a digital recording can be varispeeded while converting it back into analogue simply by running it at a different sampling-frequency, and there are a few compact disc players, R-DAT machines, and multitrack digital audio workstations which permit a small amount of such adjustment. But vari-speeding a digital *recording* can only be done on expensive specialist equipment, often by a constant fixed percentage, *not* adjustable while you are actually listening to it. Furthermore, the process results in fractional rounding-errors, as we saw earlier.

So it is vital to make every effort to get the playing-speed of an analogue medium right *before* converting it to digital. The subject is dealt with in Chapter 4; but I mention it now because it clearly forms an important part of the overall strategy. Discographical or musical experts may be needed during the copying session to select the appropriate playing-speed; it should not be done *after* digitisation.

With other processes (notably noise-reduction, Chapter 3, and equalisation, Chapters 5, 6 and 11) it may be necessary to do a close study of the relationship between the transfer and processing stages. The analogue transfer stage cannot always be considered independently of digital processing stage(s), because correct processing may be impossible in one of the domains. For readers who need to know the gory details, most digital processes are impotent to handle the relative phases introduced by analogue

equipment (section 2.11), and become impotent as zeroes or infinities are approached (especially very low or very high frequencies).

To take this thought further, how do we judge that a digital process is accurate? The usual analogue tests for frequency-response, noise, and distortion, should always be done to show up unsuspected problems if they exist. But measuring a digital noise reduction process is difficult, because no-one has yet published details of how well a process restores the original sound. It may be necessary to set up an elementary "before-and-after" experiment - taking a high-fidelity recording, deliberately adding some noise, then seeing if the process removes it while restoring the original high-fidelity sound. The results can be judged by ear, but it is even better to compare the waveforms (e.g. on a digital audio editor). But what tolerances should we aim for? This writer's current belief is that errors must always be drowned by the natural background noise of the original - I do not insist on bit-perfect reproduction - but this is sometimes difficult to establish. We can, however, take the cleaned-up version, add more background-noise, and iterate the process. Eventually we shall have a clear idea of the limitations of the digital algorithm, and work within them.

Draft standards for the measurement of digital equipment are now being published, so one hopes the digital audio engineering fraternity will be able to agree common methodology for formal engineering tests. But the above try-it-and-see process has always been the way in which operators assess things. These less-formal methods must not be ignored - especially at the leading edge of technology!

One other can-of-worms is becoming talked about as I write - the use of "neural methods." This is a buzzword based on the assumption that someone's expertise can be transferred to a digital process, or that an algorithm can adapt itself until the best results are achieved. You are of course free to disagree, but I consider such techniques are only applicable to the production of service-copies which rely on redundancies in the human hearing process. For the objective restoration of the power-bandwidth product, there can be no room for "learning." The software can only be acceptable if it always optimises the power-bandwidth product, and that is something which can (and should) be the subject of formal measurements. Ideally, any such assumptions or experiences must be documented with the final recording anyway; so when neural techniques do not even allow formal documentation, the objective character of the result cannot be sustained.

3.11 Digital recording media

Although I don't regard it of direct relevance to the theme of my book, I must warn innocent readers that conservation problems are not necessarily *solved* by converting analogue sounds to digital media. Some digital formats are more transitory than analogue formats, because they have shorter shelf-lives and less resistance to repeated use. One paper assessed the shelf life of unplayed R-DAT metal-particle tapes as 23 years, and you certainly don't want to be cloning your entire collection at 22-year intervals! Your digitisation process must therefore include careful choice of a destination medium for the encoded sounds. I could give you my current favourite, but the technology is moving so rapidly that my idea is certain to be proved wrong. (It could even be better to stick everything on R-DAT now, and delay a firm decision for 22 years).

It is also vital to store digitised sounds on media which allow individual tracks and items to be found quickly. Here we must outguess posterity, preferably without using copyright software.

I shall ask you to remember the hassles of copy-protection (section 3.8 above), and next I will state a few principles you might consider when making your decision. They are based on practical experience rather than (alleged) scientific research.

(1) It is always much easier to reproduce widely-used media than specialist media. It is still quite cheap to install machinery for reproducing Edison cylinders, because there were over a million machines sold by Edison, and both enough hardware and software survives for experience also to survive.

(2) The blank destination-media should not just have some "proof" of longevity (there are literally thousands of ways of destroying a sound recording, and nobody can test them all)! Instead, the physical principles should be understood, and then there should be no self-evident failure-mechanism which remains unexplained. (For example, an optical disc which might fade).

(3) The media should be purchased from the people who actually made them. This is (*a*) so you know for certain what you've got, and (*b*) there is no division of responsibility when it fails.

(4) Ideally the media (*not* their packaging) should have indelible batch-numbers, which should be incorporated in the cataloguing information. Then when an example fails, other records from the same batch can be isolated and an emergency recovery-programme begun.

(5) On the principle "never put all your eggs into one basket," the digital copy should be cloned onto another medium meeting principles (2) to (4), but made by a different supplier (using different chemicals if possible), and stored in a quite different place.

So, after a long time examining operational strategy, we are now free to examine the technicalities behind retrieving analogue signals from old media.

REFERENCES

1: anon, "SDMI chooses MusiCode from Aris to control Internet copying" (*news item*), London: One To One (magazine), Issue 110 (September 1999), page 10.
2: ibid, pages 73-74 and 77.
3: Barry Fox, "technology" (*article*), London: Hi-Fi News & Record Review (magazine), Vol. 44 No. 10 (October 1999), page 27.

4 Grooves and styli

4.1 Introduction

This chapter will be about "mechanical" sound recordings, in which the sound waves were translated into varying waveshapes along the length of a spiral groove. These waves are nearly always "baseband" - that is, the sound frequencies were recorded directly, instead of being modulated onto a "carrier frequency" in some way. The principal exception will be quadraphonic recordings using the CD-4 system, which I shall leave until we explore spatial sounds in section 10.16.

A century of development lies behind sound recording with mechanical techniques. The technology is unlike any other, since a frequency range of some ten octaves might be involved, with signal-to-noise ratios in the upper-sixties of decibels. Thus the power-bandwidth products of the best mechanical recordings can rival anything else analogue technology has achieved.

In this chapter I shall be considering the best ways to extract the original sound from disc or cylinder grooves. It will deal with the groove/stylus interface and also electronic techniques for minimising surface noise and distortion. Thus we will have taken the story to the point where electrical waveforms are coming out of a socket in anticipation of adjustments to playing-speed or frequency-equalisation. At this socket we will have extracted all the frequencies from the groove while minimising distortion, and will have reduced the surface noise as far as we can; thus we will have recovered the maximum power-bandwidth product. This is the fundamental limitation. After that we may refine the sound if we want to.

It is an old adage that if the groove/stylus interface is wrong, it is pointless to rely on electronics to get you out of your difficulties. I think this is something of an oversimplification, although the basic idea is correct. My point is that we should really work upon the groove/stylus interface *at the same time as* electronic noise reduction. Some electronic treatments are very good at removing showers of loud clicks, for instance. When this is done, it is then possible to choose a stylus to reduce the remaining steady hiss. Had the stylus been chosen without the de-clicker in place, it might not have been the one which gave the least hiss. This would be wrong, because the current state-of-the-art is that it is much more difficult to approach the intended sound in the presence of steady hiss.

Although I shall leave electronic treatments until the sections following 4.15, please remember the two topics should actually go hand-in-hand. And once again, I remind you of the importance of having the best possible originals, as we saw in section 2.6.

I shall be making great use of the term "thou" in this chapter. This means "thousandths of an inch", and has always been the traditional unit of measurement for styli and grooves. In North America, the term "mil" has the same meaning. Although it would be perfectly acceptable to use the metric term "microns", where one micron is one-millionth of a metre, specialist stylus-makers still speak in "thou", so I shall do the same. One thou equals 25.4 microns.

Two other terms I must define are "tracing" and "tracking." I do so because many textbooks confuse them. "Tracing" concerns the way a stylus fits into, and is modulated

by, the groove. "Tracking" concerns the alignment of the reproducer compared with the geometry of the original cutting lathe.

4.2 Basic turntable principles

I assume my readers know what a turntable is, but they may not be aware that the equivalent for cylinder records is called a "mandrel." My first words must be to encourage you to use "a good turntable (or mandrel) and pickup". It isn't easy to define this; but to recover the best power-bandwidth product, the unit must have a lower background-noise and a wider frequency-range than any of the media being played on it. This is something we can measure, and does not need any "black magic."

The unit should also have reasonably stable speed, although we shall come to a technique in section 4.16 where this need not necessarily be to the highest standards. In Chapter 4 we shall be studying the problem of establishing the correct playing-speed for rogue records. These are in the minority, but at least one machine must have variable playing-speed so this matter may be researched. (It may be necessary to have two or more machines with different features to get all the facilities).

The unit should be insulated so it is not significantly vibrated by sounds from the loudspeakers, loose floorboards, etc; this may mean fixing it to a massive mounting-plate, and suspending it in compliant supports. All this is good engineering practice, and needs no further comment from me.

Rather more difficult to define is that the unit should have "low colouration." This is much more a black magic subject. Basically it means that the wanted sounds are reproduced without "hangover" - that is, the machinery should not contribute resonances or sounds of its own which continue after the wanted sounds have stopped.

The main point about hangover is that the stylus and record themselves generate it. As the stylus is vibrated by the groove, reaction-forces develop in the disc or cylinder itself. It is important that such vibrations are quelled instantly, especially in the presence of clicks and pops. When we remove these clicks and pops using electronic techniques, we may not be able to get a clean-sounding result if hangover exists in the record itself or the pickup arm.

To deal with the record first. We cannot prevent the vibrations being set up; we can only ensure they are eliminated as quickly as possible. One method is to use a heavy rubber turntable-mat in close contact with the back of the disc. The only suitable kind of mat is the kind with a recessed area about four inches across in the middle, so the disc makes good contact irrespective of whether the label-area is raised or not. As I say, it helps if the mat is thick and heavy; the kind with lightweight ribs is pretty ineffective. Some vibrations can also be attenuated by a clamp - a heavy weight placed over the record label.

The Pink Triangle Turntable is designed without a mat at all. It is made from a plastics material of the same specific gravity as vinyl. Any vibrations set up in the vinyl are efficiently conducted away without being reflected back towards the pickup. Basically this is a good idea, but it cannot give perfect reproduction unless the disc is in intimate contact all over. You should be prepared to flatten warped records to cut down wow and tracing errors, anyway; please see Appendix 1 for details. In my experience, Pink Triangle's method is as good as the best turntable mats, but not better.

I can see no virtue in "anti-static" turntable mats, especially ones that aren't thick and heavy. A more suitable method of neutralising electrostatic charges is to keep an

open saucer of lukewarm water near the turntable, perhaps with a sponge in it for increased surface-area. Turntable lids are usually provided to keep dust off, and in my opinion they are potential sources of hangover; but in cold frosty weather (when the natural humidity is zero), they can define a microclimate where a high degree of humidity may be established. Probably the optimum solution is a removable turntable lid, perhaps with a motor-operated pickup lowering mechanism for use when the air is particularly dry and the lid must be shut.

The Pink Triangle company also make a mechanism which brings eccentric records on-centre. While it is important to play records on-centre, you may not need this if the records are your own property. Frankly, attacking the centre-hole with a round file is just as good; but after the disc has been positioned, you may need the clamp to hold it in place through the subsequent operations.

To extend the above ideas to cylinders: Edison invented the "tapered mandrel," and the vast majority of cylinders have tapered orifices so the cylinder cannot be loaded backwards. But the system has the disadvantage that the cylinder is subjected to tensile stresses as it is pushed on, which may split it. After you have done this once, you know how hard not to push! Seriously though, practicing with some sacrificial cylinders is a good way of learning.

Another problem is that playing the cylinder may generate reaction stresses between cylinder and mandrel. After a minute or two, these have the effect that the cylinder will try to "walk off the mandrel," by moving towards the narrow end in a series of very small steps. My answer is very low-tech: a rubber-band pressed up against the end of the cylinder! Eccentric or warped cylinders can often be helped with pieces of paper packed between the ends of the cylinder and the mandrel. All these disadvantages can be avoided with a machine which presses the cylinder between two contact points, holding the cylinder by its ends (Ref. 1, for example).

To reduce the overhang difficulty, I recommend a plain solid tapered mandrel made from anodised aluminium. I do not recommend anything compliant (analogous to a turntable-mat), although Reference 1 involves just that! Once again, the ideal might be two machines which provide all the features between them.

4.3 Pickups and other devices

"Pickup" is the term for a device which touches the groove as the record rotates, and translates the sound recorded in the groove into analogue electrical signals. But before we go any further, it is my duty to mention three other ways of doing the job, and why they aren't satisfactory for a sound archive. It is quite possible one of these techniques may blossom into something suitable one day. In my judgement this hasn't happened yet; but you should know the principles, so you may judge them if they improve.

The first is the original principle invented by Edison, which is to couple the stylus to something (usually a "diaphragm") which vibrates, and radiates sound directly into the air. Having stated this principle, it would normally be followed by some hundreds of pages dealing with the design of diaphragms, horns, and other acoustico-mechanical devices for improved fidelity and "amplification" ! I have put that last word in quotation-marks deliberately. The laws of conservation of energy make it clear that you cannot get "amplification" without taking energy from the rotating record, and the more you attempt this, the more wear you cause to the original record. Obviously we don't wish to wear out our records; so I propose to abandon this principle, although there will be other

lessons for us further on. The remaining principles imply electronic amplification somewhere, which does not stress the record itself.

The next principle is to play the record by looking at it with beams of light. There is a fundamental difficulty here, because the sound waves recorded in the groove may be both larger, and smaller, than typical wavelengths of light. Thus it is necessary to invent ways of looking at the groove so both large and small waves are reproduced accurately. At the time of writing the most successful of these is an analogue machine using infra-red light modulated at an ultrasonic frequency, and demodulating it using radio-reception techniques. This has three practical disadvantages over mechanical pickups. First, dirt is reproduced as faithfully as the wanted sound. Second, the frequency response is limited at the high end in a way which makes it difficult to cure the first problem. Third, the hardware can only play grooves with straight-sided walls (which we shall come to later), and only those made of something which will reflect infra-red light.

The third principle is to use an optical sensor. Here the idea is to measure the entire recorded surface in three dimensions. This might be done by scanning it with a laser-light detector at intervals of less than one micron, measuring the third dimension ("depth") by sensing when the laser light spot is in focus. This results in a vast file of some gigabytes of numerical data, which might be processed into a digital recording of the original sound.

4.4 Conventional electrical pickup considerations

We now come to the way a pickup is carried across the record to minimise geometrical sources of distortion. As this is not a textbook on conventional audio techniques, I shall not describe tracking distortion in detail, but only the specific problems encountered by present-day operators playing old discs.

It is generally assumed that discs were mastered by a cutter which moved across the disc in a straight line, whereas most pickups are mounted on an arm which carries the stylus in a curved path. In 1924 Percy Wilson did the basic geometrical study for minimising distortion from this cause (Ref. 2). This study remains valid today, but nowadays we use Wilson's formulae slightly differently. Wilson originally sought to minimise tracking error as averaged over the whole disc. Nowadays we minimise the tracking error at the inner recorded radius, which is usually taken as two and a quarter inches (56mm). There are two reasons for this: (1) the effects of tracking error are much worse at the inner radius; (2) most music ends with a loud passage, and loud passages are more vulnerable to tracking distortion.

The result of all this is that a pivoted pickup-arm should, in effect, have a bend in it so that the cartridge is rotated clockwise with respect to an imaginary straight line joining the stylus to the pivot. The exact angle varies with the arm's length, but is usually in the order of twenty degrees. In addition, the stylus should overhang the centre-pin of the turntable by an amount which also varies with arm length, but is of the order of about 15mm. When a pickup arm is installed in this manner, minimum tracking distortion is assured from conventional records. But operators should be aware that the "alignment protractor" supplied with many pickup arms will *not* give the correct alignment for *un*conventional records.

A pivoted arm is much more amenable to experimental work than a non-pivoted one such as a parallel tracking arm. This doesn't mean that either type is inherently superior, only that one must use the right tool for the job. Many pivoted arms have an oval-shaped base for the actual pivot, and the whole arm can be moved bodily towards or

away from the turntable centre. This enables you to neutralise tracking distortion when the disc ends at an unusual radius. Coarsegroove 33rpm discs, for example, may finish 100mm from the centre; but tracking-distortion can be quite noticeable in these circumstances because the sound waves are packed close together in a coarse groove. The arm can be moved towards the turntable slightly to make the cartridge perfectly tangential to the inner radius. On the other hand, many small 78rpm discs of the late 1920s and early 1930s were recorded much further in; the worst example I know ends only 20mm from the centre hole. The tracking distortion is terrible under these conditions, and may be reduced by moving the whole arm a centimetre or so away from the turntable centre.

However, tracking distortion can be totally eliminated from conventional records by means of a "parallel tracking arm" - a mechanism which carries the pickup across the disc in a straight line. In practice this is difficult to achieve without causing other problems, so parallel tracking arms are more expensive; but in this situation, the centre-line of the cartridge should be aligned perpendicular to the direction of motion, and the path of the stylus should pass through the centre-pin of the turntable.

However, I must report that, although a parallel tracking arm eliminates tracking distortion on conventional records, it is practically impossible to do anything about *un*conventional ones. In practice, these fall into two types. (1) Discs cut on a lathe where the cutterhead was correctly aligned, but the cutting stylus was inserted askew. In this case the tracking error is the same at all radii. (2) Discs cut on a lathe which carried the cutter in a straight line *not* passing through the centre of the disc, but along a line parallel to the radius; or, what comes to the same thing, discs cut on a lathe whose cutterhead was not mounted squarely. In these cases the tracking error varies with radius. These features result in various types of distortion which we shall examine in detail later.

Whether it is a pivoted arm or a parallel-tracker, the arm itself should not contribute any significant resonances after being shock-excited by cracks and bumps. You may need a long arm (such as the SME 3012) for playing outsized discs; but any long arm will have noticeable resonances. Since the original Model 3012 was made, SME have an upgrade-kit comprising a trough of silicone damping fluid. This greatly reduces the resonances, but a later design (such as their Series V) is preferable for conventional-sized discs. All other things being equal, a parallel-tracker has less metal which can vibrate; experience with massive de-clicking operations tends to show that this type is better for badly cracked and scratched records.

All cylinders were mastered with a straight-line movement parallel to the axis, so this geometry should be followed in principle. All other things being equal, it is of course irrelevant whether it is actually the pickup or the cylinder which moves; but other considerations (such as groove-jumping) may force one method or the other. A machine called "Ole Tobias" at the National Library of Norway combines the principle of having no mandrel as mentioned at the end of 4.2, with a pivoted tonearm whose pivot is driven in a straight line parallel to the axis. This would seem to combine all the advantages of both; but I have no knowledge of the time it takes to get each cylinder running concentrically.

4.5 Operational procedure for selecting a stylus

Because there are techniques for dealing with crackle and clicks, the maximum power-bandwidth product comes from choosing a record with the best balance between the basic hiss and the distortion due to wear. Although psychoacoustic tricks exist for reducing hiss, there are currently no cures for wear, so priority should be given to an unworn copy.

An experienced transfer operator will be able to choose the correct stylus simply by looking at the grooves. I am afraid this is practically impossible to teach; the operator has to look at both groove walls, the groove bottom, and the "horns" (if any; see section 4.6). A "point source" of light is needed (*not* fluorescent tubes), preferably in an "Anglepoise" so you may choose different ways of looking at the surface. For selecting a stylus, put the anglepoise behind your shoulder so you are looking down the beam of light; then turn a disc about a *horizontal* axis between the two hands, while watching how the overall amount of reflected light varies with the angle of the disc. I know that's a totally inadequate explanation, but I simply can't do any better. Until you have learnt the trick, you will be obliged to go through a procedure of consecutive feedback-loops to identify the best stylus.

Thus you may have to compare two or more styli to see which gives the greater power-bandwidth product. Unfortunately practical pickup cartridges cannot withstand frequent stylus-changes (which in many cases can only be done by the manufacturer anyway). So we must use exchangeable headshells, which will annoy the hi-fi buffs. Allow me to deal with this objection first. Exchangeable headshells are inherently heavier than fixed headshells. But the reduction of mass is only significant when dealing with warped or eccentric records at very low playing forces; various types of distortion can occur if there is any tendency for the pickup to move up and down or to and fro. Frankly, it is much better to have flat concentric discs to start with! For an archive copy this is essential anyway, as it is the best way to minimise speed inconsistencies.

So the professional transfer operator will have several cartridges mounted in headshells ready for comparison. Sometimes, however, we find ourselves at the point of diminishing returns. When we have got reasonably sensible noises out of the groove, it may require a lot of work to make a not-very-significant improvement to the power-bandwidth product. By the time I have unscrewed one head-shell and tried another, I find I have forgotten what the first one sounded like. There are two cures: (1) Transfer one version before changing the shell; (2) to have two pickup arms playing the same record and switching between them (this is a better way, as it reduces wear-and-tear on the headshell contacts).

To close the feedback loop, and expedite the choice of one stylus from dozens of possibilities, we must learn the mechanisms involved and their effects upon the reproduction. This chapter therefore continues with a look at the history of grooves and styli. Whenever we come across a technique which is still applicable today, I shall interrupt the history lesson and examine the technique in more detail. I am afraid this will mean a rather zig-zag course for my argument, but I hope that sub-headings will allow you to concentrate upon one strand or the other if you wish.

4.6 U-shaped and V-shaped grooves

I shall talk mainly about two kinds of groove – "U-shaped" and "V-shaped" - but I shall not formally define these terms. I use them to differentiate between two philosophies for playback purposes; you should not assume that all "V-shaped" grooves have straight walls and sharp bottoms, for example. And recordings made at the dawn of sound recording history do not fit either category.

Edison's tinfoil phonograph did not cut grooves. It indented them in a sheet of metal commonly known as "tinfoil." The noise-level of the groove was determined principally by the physical properties of the foil. It was virtually impossible to remove it and replace it correctly without either corrupting the indentations or crinkling the sheet; and there was inevitably a once-per-revolution clunk as the stylus crossed the seam where the foil was wrapped round the mandrel. These features were largely responsible for the eclipse of the phonograph as a practical recording machine during the years 1878 to 1887. They also explain why so few tinfoils survive today, and those in unplayable condition.

Bell and Tainter's "Graphophone" circumvented these difficulties by using pre-shaped cardboard cylinders coated in a wax-like substance called "ozokerite." Thus the problems of coiling up the tinfoil, aligning it on the mandrel, and arranging for an inoffensive seam, were avoided. But Bell & Tainter's fundamental improvement was that the groove was *cut* instead of indented. The recording machine was fitted with a stylus which actually removed a continuous thread of ozokerite, leaving behind a fine clean groove with much lower noise. (In parentheses, I add that the Graphophone used an *untapered* mandrel. So there may be ambiguity about which is the start and which is the end of the recording).

Edison's "Improved Phonograph" of 1888 adopted the cutting idea, but he favoured cylinders made of solid wax much thicker than the Graphophone's layer of ozokerite. It was therefore possible to erase a recording by shaving it off. This was much better suited for dictation purposes, which is how both the Graphophone and the Improved Phonograph were first marketed. I do not know the details of Graphophone cutters, but I do know that Edison's Improved Phonograph used a sapphire cutting-tool. Sapphire is a jewel with a hardness greater than any metal. Anything less hard was found to wear out quickly. This was the main reason behind the commercial failure of the Graphophone, because a blunt cutter would not make a quiet groove.

Artificial sapphires were made for the jewelled bearings of watches. They were cylindrical in form and smaller than a grain of rice, about one-hundredth of an inch in diameter. To make a phonograph cutter, one end was ground flat and mounted so it would dig end-on into the rotating wax. The sharp edge where the flat end met the curved rim would be where the swarf was separated from the cylinder, leaving behind a groove so smooth it would reflect light. In practice, the cutter would be tilted at a slight angle, and the front face ground to a complimentary angle. This left a groove bottom which wasn't shaped like an arc of a circle, but an arc of an ellipse with a low degree of eccentricity. This is what I mean when I talk about "U-shaped" grooves.

In the case of Edison machines, recordings were reproduced by another jewel, this one deliberately "blunt" so it would not cut the wax again, but small enough to run along the bottom of the groove. The vertically-modulated sound waves would cause the reproducing stylus to be vibrated up and down as it pressed against the groove bottom, and thus the sound would be extracted. Later developments resulted in playback styli made to a specific diameter to fit the grooves, minimising noise and wear. Edison

established standards for his two-minute cylinders, his four-minute cylinders, his "Voicewriter" dictation-machines, and his "Diamond" discs. Edison also showed that minimum distortion occurred with a button-shaped playback stylus (the correct geometrical term is an oblate spheroid). This was designed to sit across a plain groove whilst remaining in contact all round, while its minor radius was sufficiently small to follow the most intricate details of the recorded waveform.

Meanwhile, back in 1888, Emile Berliner was developing a quite different way of recording sound. There were three fundamental differences. (1) He preferred discs to cylinders, which gave him two advantages. His reproducing machines needed no mechanism to propel the reproducing stylus, the disc itself would do it; and he could mass-produce copies of his records like printing. (2) His styli vibrated side-to-side rather than up-and-down. The groove walls therefore pushed the playback styli to and fro rather than the unidirectional propulsion of the hill-and-dale (vertical cut) format. (3) He did not cut grooves, but used an acid-etching process.

"Acid-etched" disc recordings, made between 1888 and 1901, therefore have grooves of rather indeterminate cross-section. Partly because of this, and partly because Berliner was competing with cylinder manufacturers on cost grounds, Berliner used relatively soft steel reproducing needles and made his discs in relatively abrasive material. The first few seconds of groove would grind down the tip of the reproducing needle until it had its maximum area of contact, thereby ensuring the needle would be propelled by the groove walls, while his machines avoided the cost of jewelled playback styli. On the other hand, steel needles could only be used once; and this philosophy remained the norm until the 1950s.

In 1901 Eldridge Johnson (founder of the Victor Company) adapted the wax-cutting process to the mastering of disc pressings, so the groove now had consistent cross-section throughout the side of the disc. For several decades they were usually U-bottomed like hill-and-dale recordings. Although the abrasive nature of the pressings did much to hide the advantages, the wax masters and the stampers had smoother surfaces than acid-etched recordings, and today much of our restoration work consists of trying to get back to the low noise-level of the wax masters.

The vast majority of such pressed records were played with steel needles. The only exceptions were collections belonging to wealthier or more careful collectors, who used "fibres" (see section 4.8).

In 1911 a British inventor, P. J. Packman, patented a new type of cutting stylus in which a cylindrical sapphire had its axis perpendicular to the wax, rather than substantially parallel to it. (Ref. 2). His aim was to cut deeper grooves. He wanted to pack more sound into a given space, and reasoned that if one used hill-and-dale recording, one would not have to leave space between the grooves for lateral modulation. By combining hill-and-dale recording with a finer groove-pitch and the technique of an abrasive record to grind a steel needle, he hoped to make inexpensive long-playing disc records; a couple of hundred were actually published under the tradename "Marathon." They were not a success; however, the principle of Packman's cutter was gradually adopted by the rest of the sound recording industry.

There were several advantages to a relatively deep groove. The deeper it was, the less likely the sound would be corrupted by scratches and dirt. Also a reproducing stylus was less likely to "skid," or to be thrown out of the groove by heavy modulation. These advantages meant it was easier to accommodate louder sounds. There could be a greater area of contact between stylus and groove, so there could also be less hiss as we shall see in section 4.8.

If one tries to cut a groove of U-shaped cross-section which is impracticably deep, the walls will become nearly vertical at the surface of the disc. A number of difficulties come to light if this happens. During the cutting process, the swarf does not separate cleanly, because material being forced up from the bottom of the groove causes shearing action (rather than cutting action) at the top. Even if this problem were overcome, it would be much more difficult to press or mould records from a negative with ridges of near-semicircular shape. The material would adhere to the stamper rather than separate cleanly, because of different coefficients of thermal contraction as stamper and material cooled. When the groove walls are less than 45 degrees, the thermal contraction results in the record being pushed away from the stamper; when it is greater than forty-five degrees, the record tends to be gripped by the stamper.

Therefore the deepest possible groove can only be a V-shape rather than a U-shape, with no part of the groove walls greater than forty-five degrees from the horizontal. This therefore represents the ultimate practicable groove-shape to make use of the advantages I have just described.

Nowadays, when most people have done applied mathematics at school, the idea of a force being resolved into two components is commonplace. But evidently this wasn't the case in the first quarter of this century; it was thought grooves must have flat bottoms if the record was to take the playing-weight of acoustic soundboxes (over a hundred grams). Today we know that a V-shaped groove is equally capable of bearing such a weight, since the force exerted upon each inclined groove wall can be resolved into horizontal and vertical components, and the two horizontal components from each of the two walls cancel. Packman's groove worked this way, although he did not claim it as part of his invention. During the first World War the English Columbia company adopted V-shaped grooves for its records, I suspect largely because they had much worse surface noise than their competitors at that time. But the mistaken idea of U-bottomed grooves being inherently better remained the dominant philosophy until the early 1930s.

What forced the change was the advent of the auto-changer for playing a stack of records without human intervention. Reproducing needles suddenly had to be capable of playing eight consecutive sides. Less wealthy customers still used relatively soft steel needles, so the records had to retain abrasive qualities to grind them until there was a perfect fit - an early case of "downwards compatibility." Only the very hardest stylus materials would stand up to eight abrasive sides, and various forms of tungsten were tried, followed eventually by the renaissance of jewels. To ensure the grooves would always propel such styli backwards and forwards in the lateral plane, the *walls* had to be in control. This was impossible so long as different records had U-bottomed grooves of different sizes and the playback styli couldn't adapt themselves. So the industry gradually changed to V-shaped grooves cut by Packman-type cutters, a process which was complete by 1945.

Although Packman's patent shows a V-shaped cutter coming to a definite point, sapphires of this design are very fragile. Sapphire is very hard and it resists compression tolerably well, but it has little shear strength. Cutting a V-shaped groove with a sharp bottom is practically impossible. Instead, the tip of the cutter is deliberately rounded, and the resulting groove actually has a finite radius in its bottom. In 78rpm days this might be anything between 1 thou and 2.5 thou, even in nominally V-shaped grooves. With the introduction of microgroove, the bottom radius might be 0.3 to 0.7 thou. If we consider mass-produced pressings, the radius tended to increase as the stamper wore, and greater radii may be encountered in practice.

Before it became possible to copy a disc record electrically (in about 1930), a factory might "restore" a negative by polishing it, so the pressing would have a groove with a flat (or flattish) bottom, no matter what the original groove shape was. This was done to clear up background-noise due to irregularities in the bottom of the groove, which were reproduced loud and clear when played with a steel needle ground to fit. Background-noise was certainly ameliorated, but the process was not without side-effects. A steel needle would take longer to grind down, resulting in an extended period of wear; and before modern stylus-shapes became available, such blunter styli could not trace the high frequency detail.

To continue my history lesson: Cecil Watts invented the cellulose nitrate lacquer recording blank in 1934. This comprised a layer of lacquer upon a sheet aluminium base, and although there were many alterations in the detailed composition of the lacquer and in the material used for the base, as far as I know cellulose nitrate was always the principal constituent. His development accompanied rivals such as "gelatine", "Simplat," "Permarec," and others, but these were all aimed at amateur markets. Only "nitrate" was adopted by professionals, because (when new) it had lower background-noise than any other analogue medium, either before or since. (For some reason it was called "acetate" for short, although as far as I know there was never a formulation relying on cellulose acetate as its principal component. I shall call it "nitrate.")

Nitrate was gradually adopted by the whole disc-recording industry to replace wax, which was more expensive and too soft to be played back; the changeover was complete by 1950. Wax cutters had had a plain sharp cutting edge (known, for some reason, as a "feather edge.") Packman-type sapphires with feather-edges could not withstand the extra shear stresses of nitrate, so steel cutters were widely used for such discs. However it was noticed they sometimes gave *lower* surface noise as they wore. There was a great deal of hocus-pocus pronounced about this subject, until the New York cutting stylus manufacturer Isabel Capps provided the correct explanation. The swarf was being separated by the front face of the cutter, as intended; but when it was slightly blunt, the following metal pushed the groove walls further apart and imparted a polishing action. When this was replicated by adding a "polishing bevel" to a sapphire cutter, there was a major improvement in background-noise and the sapphire was better able to withstand the shear stresses at the same time. From the early 1940s sapphire cutters with polishing-bevels became normal for cellulose nitrate mastering.

These polishing-bevels had the effect of increasing the minimum possible recorded wavelength. Although insignificant compared with the losses of playing contemporary grooves with contemporary styli, they caused a definite limit to the high-frequency reproduction possible today.

Because the polishing-bevel pushed some of the nitrate aside, the result was that miniature ridges were formed along the top edge of the groove walls, called "horns." If not polished off the "positive," they are reproduced upon the pressed records, and when you have learnt the trick of looking at them, the horns provide conclusive evidence that nitrate was used rather than wax. We may need to know this when we get to section 4.11.

With microgroove recording it was necessary to adopt another technique to allow small recorded wavelengths. The cutting stylus was heated by a red-hot coil of wire. The actual temperature at the cutting-edge proved impossible to measure in the presence of the swarf-removal suction, but it was literally like playing with fire. Cellulose nitrate is highly inflammable, and engineers have an endless supply of anecdotes about the resulting conflagrations. By almost melting the lacquer at the cutting-edge, it was found

possible to make the polishing-bevels much smaller and improve the high frequencies, to reduce the background-noise of the master-lacquer, and to extend the cutter's life at the same time. (Ref. 3).

The final development occurred in 1981-2, when "direct metal mastering" was invented by Telefunken. To oversimplify somewhat, this involved cutting a groove directly into a sheet of copper. A diamond cutter was needed for this, and for a number of reasons it was necessary to emulate the polishing action by an ultrasonic vibration of the cutter. A decade later, more than half the disc-cutting industry was using the process.

This has been a grossly oversimplified account of what became a very high-tech process; but I mention it because operators must often recognise techniques used for the master-disc to ensure correct geometry for playback.

4.7 The principle of minimising groove hiss

We now come to the problems of reproducing sound from grooves with fidelity. This is by no means a static science, and I anticipate there will be numerous developments in the next few years. The power-bandwidth principle, however, shows us the route, and quantitatively how far along the road we are. Most of what I shall say is applicable to all grooved media; but to save terminological circumlocutions, I shall assume we are trying to play a mono, lateral-cut, edge-start shellac disc, unless I say otherwise.

The irregularities in the walls of the groove cause hiss. These irregularities may be individual molecules of PVC in the case of the very best vinyl LPs, ranging up to much larger elements such as grains of slate-dust which formed a major constituent of early 78s. The hiss is always the ultimate limit beyond which we cannot go on a single copy, so we must make every effort to eliminate it at source. In fact, steady hiss is not usually the most noticeable problem, but rather crackles and pops; but we shall see later that there are ways of tackling those. It is the basic hiss that forms the boundary to what is possible.

The only way to reduce the basic hiss from a single disc is to collect the sound from as much of the modulated groove walls as we can. It is rather like two boats on a choppy sea; a dinghy will be tossed about by the waves, while a liner will barely respond to them. Playing as much of the groove as possible will imitate the action of a liner. We can quantify the effect. If our stylus touches the groove with a certain contact area, and we re-design the stylus or otherwise alter things so there is now twice the area of contact, the individual molecules or elements of slate dust will have half their original effect. In fact, the hiss will reduce by three decibels. So, *if* the basic hiss is the problem, we can reduce it by playing as much of the groove as we possibly can.

Please note that last caveat – "*If* the basic hiss is the problem." That is an important point. If the noise we hear is *not* due to the basic structural grain of the record, this rule will not apply. Suppose instead that the record has been scratched at some time in the past, and this scratch has left relatively large protruding lumps of material in the walls of the groove. If we now attempt to double the contact area, the effects of the scratch will not be diluted; to the first approximation, the stylus will continue to be driven by the protruding lumps, and will not be in contact with the basic groove structure. Thus the effects of the scratch will be reproduced exactly as before. To use our boat analogy again, *both* the liner and the dinghy will be equally affected by a tidal wave.

So whatever we subsequently do about the scratches and clicks, our system must be capable of playing as much of the groove as possible, in order to reduce the basic hiss of an undamaged disc. I shall now consider some ways of achieving this ideal - which, I

must repeat, is not necessarily an ideal we should always aim at, because of other sources of trouble.

4.8 "Soft" replay styli

I must start by making it quite clear that there are several quite different approaches we might take. Nowadays we tend instinctively to think in terms of playing a mono lateral-cut edge-start shellac disc with an electrical pickup fitted with a jewel stylus, but it is only right that I should first describe other ways of doing it. The Nimbus record company's "Prima Voce" reissues of 78s on compact discs were transferred from an acoustic gramophone using bamboo needles, and whatever your opinion might have been about the technique, you could only admire the lack of surface noise.

For modern operators used to diamonds and sapphires, it is necessary for me to explain the thinking behind this idea. I use the unofficial term "soft" needle for any stylus which is softer than the record material. It forms a collective term for the needles better known as "bamboo", "thorn", or "fibre", but please do not confuse my term with so-called "soft-*toned*" needles; I am referring to physical softness.

Most shellac lateral-cut discs were deliberately designed to be abrasive, because they had to be capable of grinding down a steel needle. Microscopic examination would then show that the grooves were populated with iron filings embedded in the walls; the result was additional surface noise. From about 1909, quality-conscious collectors used soft needles instead. (Ref. 4). They were obtained from various plants and treated in various ways, but all worked on the same principle. A slice from the outer skin of a bamboo, for example, was cut triangular in cross-section. Soundboxes with a triangular needle-socket were obtainable. The needle could be re-sharpened by a straight diagonal cut; you could do this with a razor, although a special hand-tool was easier to use. This action left a point sharp enough to fit any record groove. "Bamboos" were favoured for acoustic soundboxes. The smaller "thorn" needles for lightweight electrical pickups had to be sandpapered or otherwise ground to a conical point. "Fibre" needles seems to be a collective term for the two types. All such materials had much more hardness along the grain of the fibre, and it was possible to achieve a really sharp point - as little as 0.5 thou was often achieved. (Ref. 5). Sometimes specialist dealers provided needles impregnated with various chemicals, and much effort was expended in getting the optimum balance between lubricant (when the needle was relatively soft) and dryness (when the needle was much harder, transmitted more high frequencies, and lasted longer in the groove).

At the outside edge of a shellac disc, a "soft" needle wore very rapidly until it was a perfect fit for the groove - this happened typically within one revolution, so the parts of the groove containing the music were not affected by wear. After that, the close fit with the groove-shape minimised the wear and the hiss, as we saw in our boat analogy. By having a very fine point with an acute included angle (semi-included angles of only ten degrees were aimed for), the shape would fit the groove perfectly in the direction across the groove, but would be relatively short along the groove. This was found empirically to give less distortion. I am not aware that a printed explanation was ever given, although clearly users were emulating Edison's oblate spheroid of four decades earlier, and the "elliptical" jewels of four decades later.

The hiss was also attenuated by the compliance of the needle, which however affected the reproduction of wanted high frequencies (Ref. 6). However, greater compliance meant less strain between needle and groove, so less wear at high frequencies

- exactly where steel and jewel styli had problems. Modulation of high amplitude would sometimes cause the point to shear off a "soft" needle, but collectors considered this was a small price to pay for having a record which would not wear out. Only one playing with steel would damage the record, adding crackle, and would render further fibres useless by chipping at the ends of the bundle of fibrous material. Collectors therefore jealously guarded their collections, and did not allow records out of their hands in case the fibred records became corrupted.

"Soft" needles were in common use through the 1920s and 1930s, and the pages of *The Gramophone* magazine were filled with debates about the relative merits. However these debates always concentrated upon the subjective effect; there is nothing objective, and nowadays we find it difficult to tell which gives the optimum power-bandwidth product. Soft needles were even tried on microgroove records in the early 1950s, evidently with some success (Ref. 7).

The motivation was, of course, to make contact with *all* the groove so as to reduce the hiss, as we saw earlier. Reference 6 quite correctly described the disadvantages of thorn needles (the attenuation of high frequencies and the risk that the tip might shear off), but it perversely did not say what the advantages were. To archivists nowadays, those two disadvantages are less important. We can equalise frequency response aberrations (so long as they are consistent), and we can resharpen the stylus whenever needed (as Nimbus did in the middle of playing their records, editing the resulting sections together).

A number of experiments by myself and others may be summarised as follows. A "soft" needle undoubtedly gives lower surface noise than any other, although the differences are less conspicuous when the high-frequency losses due to the needle's compliance are equalised. (The response starts to fall at 3 or 4 kiloHertz when a bamboo needle is used in an acoustic soundbox). This is apparently because, with "hard" styli, in most cases we are *not* playing the basic hiss of the record, but the damage; this does *not* mean reduced noise. But damage shears bits off a "soft" point, and is then reproduced more quietly - a sort of mechanical equivalent of the electronic "peak clipper." The reproduction is particularly congested at the end of the disc, because the needle has a long area of contact along the groove. Scientific measurements of frequency response and distortion give inconsistent results, because the tip is constantly changing its shape; thus I must regretfully conclude that a soft needle is not worthy of a sound archive. Also the process requires a great deal of labour and attention. However, it shows that better results are possible, and we must try to emulate the success of the soft needle with modern technology.

There seems to be only one case where the soft needle is worth trying - when the groove is of indeterminate shape for some reason - perhaps an etched "Berliner" or a damaged cutter. Grooves like this sometimes turn up on nitrate discs, and are not unknown in the world of pressings; operators derisively describe them with the graphic expression "W-shaped grooves." Obviously, if the groove is indeed W-shaped, or otherwise has a shape where a conventional stylus is useless, then a soft needle is worth trying.

I should like to conclude this topic by advising potential users to experiment. It is comparatively easy to construct soft needles to one's own specification, and there seems to be little harm to be caused to shellac 78s. I should also like to see a stereo pickup which takes such needles; in section 0 we shall be examining another method of reducing noise which depends upon the record being played with a stereo pickup. But I should like to remind you that if the needle develops a "flat" (which means it has a long area of contact

in the dimension along the groove), various forms of distortion become apparent on sounds which cause sharp curves in the groove. So if you are doing any serious experimenting, I recommend you make use of an Intermodulation-Distortion Test Disc. In America the obvious choice is RCA Victor 12-5-39, and in Britain EMI JH138. References 8 and 9 give some hints on how to use them and the results which may be expected.

4.9 "Hard" replay styli

I count steel needles as being neither "soft" nor "hard." They were soft enough to fit the groove after a few turns when played with pickups whose downwards force was measured in ounces, but the knife-edges worn onto the tip during this process cut into the waveform and caused distortion. They are not used today for this reason, and I can say with confidence that sacrificial trials of steel needles upon shellac discs do not give a better power-bandwidth product. If anyone wants to confirm my experiment, I should mention that "soft-toned" steel needles (soft in *volume*, that is) have extra compliance. This gives high-frequency losses like fibres, which must be equalised for a fair trial.

On the contrary, from the end of the second World War onwards, it has been considered best practice to use hard styli, by which I mean styli significantly harder than the record material. Styli could be made of sapphire, ruby, or diamond; but I shall assume diamond from now on, because sapphires and rubies suffer appreciable amounts of wear when playing abrasive 78s, and do not last very long on vinyl. The cost of a stylus (especially a specialist shape) is now such that better value is obtained by going for diamond right from the start. In the author's experience there are two other advantages. Diamonds are less likely to be shattered by the shocks imparted by a cracked disc. Also diamonds can play embossed aluminium discs; sapphire is a crystalline type of aluminium oxide, which can form an affinity with the sheet aluminium with mutual destruction.

At this point I will insert a paragraph to point out the difficulties of worn "hard" styli. The wear changes a rounded, and therefore inherently "blunt", shape, into something with a cutting edge. In practice, U-bottomed grooves cause a worn patch whose shape approaches that of the curved surface of a cylinder; V-bottomed grooves cause two flats with plane surfaces. Where these geometrical features intersect the original curved surface of the stylus, we get an "edge" - a place where the stylus has a line separating two surfaces, rather than a curved (blunt) surface. This can (and does) cut into record grooves, particularly where the groove contains sharp deviations from the line of an unmodulated spiral. Thus we cause distortion and noise on loud notes. The damage is irreversible. Therefore it is better to use diamond tips, which develop such cutting edges less easily. The increased cost is greatly outweighed by the increased life. Inspection with a 100-diameter microscope and suitable illumination is sufficient to show the development of flats before they become destructive.

The "bluntest" shape, and therefore the one least likely to cause wear, is the sphere. "Spherical" tips were the norm from about 1945 to 1963. The spherical shape was ground onto the apex of a substantially-conical jewel mounted onto the armature or cantilever of an electrical pickup. Being spherical, there were relatively few problems with alignment to minimise tracing distortion (section 4.10), so long as the cantilever was pointing in about the right direction and there was no restriction in its movement. The spherical shape gave the maximum area of contact for a given playing-weight. So acceptable signal-to-noise ratio and reasonable wear was achieved, even upon the earliest vinyl LPs with pickups requiring a downward force of six grams or more.

From 1953 to 1959, the British Standards Institution even went so far as to recommend standardised sizes of 2.5 thou radius for coarsegroove records and 1 thou for microgroove records. This was supposed to ensure disc-cutting engineers kept their modulation within suitable limits for consumers with such styli; it did not directly influence the dimensions of grooves themselves. However, the idea had some difficulties, and neither recommendation lasted long. You may often find older records needing larger styli (particularly microgroove from the Soviet Union). Something larger than 1 thou will be absolutely vital for undistorted sound here.

But first I shall mention the difficulties for coarsegroove records. We saw in section 4.6 that manufacturers were forced to change to V-shaped grooves to enable harder needles to last in autochangers. 2.5-thou spherical styli could be relied upon to play such V-shaped grooves successfully, but they were often troublesome with older U-shaped grooves. When you think about it, a hard spherical tip is very likely to misbehave in a U-shaped groove. If it is fractionally too small, it will run along the bottom of the groove and not be propelled by the laterally-modulated groove walls at all; the result is noisy reproduction and distortion at *low* levels. If it is fractionally too large, it will not go properly into the groove at all, but sit across the top edges. The result, again, is noisy; but this time the distortion occurs on loud notes and high-frequency passages. As long as customers could only buy spherical tips, the only way forward was to buy tips of different diameters. Enthusiasts had a range of spherical-tipped styli such as 2.5-thou, 3-thou, 3.5-thou, and 4-thou, specifically for playing old U-shaped grooves. We saw in section 4.6 that U-shaped grooves had, in fact, the cross-section of a ellipse with a low degree of eccentricity. It was found that a range of styli with differences of 0.5 thou, could trace nearly all such grooves. It was also found that for modern coarsegroove discs with V-shaped grooves and high levels of modulation (or high frequencies), smaller tips were helpful, such as 2-thou and 1.5-thou.

For microgroove discs, the recommended 1-thou tip was found unsatisfactory for a different reason. Pickup development was rapid, aided by Professor Hunt's papers about the effects of pickup design upon record-wear. These showed the advantages of high compliance, high areas of contact, and low effective tip-mass. Below certain limits in these parameters, Professor Hunt showed that pickups would work within the elastic limits of vinyl and cause *no* permanent wear (although instantaneous distortions could still occur). (Refs. 11 and 12). Great efforts were made to approach the ideals laid down by Professor Hunt, which the introduction of stereo and the high volumes of pop discs did little to hinder. By the end of the 1950s it was apparent that greater fidelity could be achieved with smaller spherical tips, 0.7 thou or 0.5 thou. Although such styli often "bottomed" on older records, they could trace the finer, high-frequency, details of newer discs with less distortion; thus more power-bandwidth product was recovered. There was increased disc wear due to the smaller area of contact, but this was soon reduced by improvements in compliance and effective tip-mass.

There was another consideration in the days of mono, which is of less importance now. Consider a spherical stylus playing a lateral-cut mono groove with loud modulation. Where the groove is slewed, its cross-section also narrows. To overcome this, the stylus must also rise and fall twice a cycle - in other words, it must possess vertical compliance. Even if we are only interested in the horizontal movement, the centre of the spherical stylus tip does not run exactly along the centre-line of the groove. The resulting distortion is called "pinch effect distortion," and permanent deformation of a groove was caused if a stylus had low vertical compliance.

For years two sources of distortion tended to cancel each other. The harmonic distortion which resulted from the large tip being unable to trace the fine detail was almost exactly opposite to the harmonic distortion caused when a massive stylus modified its course by deforming the groove wall. (Ref. 13). The fact that two types of distortion neutralised each other was often used deliberately, but at the cost of permanent damage to the groove. This explains why so many "pop singles" could be recorded at such high volumes. If you don't have a copy in unplayed condition, your only chances of getting undistorted sound are to use a large tip with low vertical compliance, or find another version of the same song.

4.10 Stereo techniques

To carry a stereo recording, a disc has to be capable of holding two channels of sound in accurate synchronism. Several different methods were tried at different dates. In the 1930s both Blumlein in Britain and the Bell Labs engineers in America tried cutting the left-hand sound as lateral modulation and the right-hand channel as vertical modulation. This worked, but the two systems have different distortion characteristics. So the reproduced sound had asymmetrical distortion, which was very noticeable because this cannot occur in nature. Arnold Sugden of Yorkshire England attempted the same thing using microgroove in the years 1955-6 (Ref. 14). Meanwhile Cook Laboratories in America recorded stereo by using two lateral cutters at different radii, an idea taken up by Audio Collector and Atlantic. It was found that the inevitable tracking errors of pivoted pickup arms caused small time-shifts between the two channels, which were very noticeable on stereo images. Back in the UK, Decca tried an ultrasonic carrier system. One channel was recorded using conventional lateral recording, while the other was modulated onto an ultrasonic 28kHz carrier-wave, also cut laterally.

But the solution ultimately adopted was one originally patented by Blumlein, although not actually used by him as far as I know: to record the *sum* of the two channels laterally, and their *difference* vertically. Not only do the two channels have symmetrical distortion characteristics, but the record has the advantage of "downwards compatibility," so a mono record will reproduce in double-mono when played with a stereo cartridge. This is geometrically equivalent to having one channel modulated at an angle of 45 degrees on one wall of a V-shaped groove, and the other at right-angles to it upon the other groove-wall. The convention adopted was that the wall facing the centre of the disc should carry the right hand channel, and the one facing away from the centre should have the left-hand channel. This standard was agreed internationally and very rapidly in April 1958, and I shall be assuming it from now on. The other (very rare) systems amount to "incunabula."

V-shaped grooves were standard by now, as we have seen. There were no immediate consequences to the design of hard styli, except to accelerate the trend towards 0.5 thou sphericals and ellipticals (which form the topic of the next section) as part of the general upgrading process.

But a new source of distortion rapidly became noticeable – "vertical tracking distortion." Cutters in the original Westrex 3A stereo cutterhead, and its successor the model 3B, were mounted on a cantilever whose pivot was above the surface of the master-disc. So when "vertical" modulation was intimated, it wasn't actually vertical at all; it was at a distinct angle. In those Westrex cutterheads the angle of the cantilever was twenty-three degrees, while in another contemporary cutterhead (the Teldec) there was

no cantilever at all, so when the cutter was meant to be moving vertically, it *was* moving vertically.

So besides the various tracing and tracking distortions known from lateral records, there was now a new source of trouble. When a perfect vertical sine wave is traced by a cantilever, the sine wave is traced askew, resulting in measurable and noticeable distortion. This is exactly analogous to the "tracking distortion" on lateral modulation when you play the groove with a pivoted arm (section 4.4), so the phenomenon is called "vertical tracking distortion." The solution is to ensure that cutter cantilevers and pickup cantilevers operate at the same angle. It proved impossible to design a rugged stereo pickup *without* a cantilever, so the angle could *not* be vertical. Some years of research followed, and in the meantime non-standard stereo LPs continued to be issued; but the end of the story was that a vertical tracking angle of fifteen degrees was recommended. (Ref. 15).

The first difficulty was that the actual physical angle of the cantilever is not relevant. What is important is the angle between the tip of the stylus and the point (often an ill-defined point) where the other end of the cantilever was pivoted. Furthermore, variations in playing-weight and flexing of the cantilever at audio frequencies had an effect. All this took some time to work out, and it was only from about 1964 onwards that all the factors were understood, and a pickup could be guaranteed to have a fifteen-degree vertical tracking angle at high frequencies (which is where the worst of the trouble was). Unfortunately, by 1980 a gradual drift had become apparent among pickup makers, if not disc cutting engineers; the average angle was over twenty degrees.

Similar problems applied to cutting the master-disc. Some designs of cutter proved impossible to tilt at the required angle. And more research was needed because of a phenomenon known as "lacquer-springback." We saw in section 4.6 that cellulose nitrate lacquer discs gradually took over for disc mastering, the changeover years being roughly 1936 to 1948. It was found that elasticity of the lacquer also caused some vertical tracking error, because after the cutter removed a deep section of groove, the lacquer tended to creep back under its own elasticity. This effect had not been noticed before, because the springback was consistent for lateral cutting (with constant groove depth), and for wax (which was not elastic). But the vertical tracking error from lacquer alone might be twenty degrees or so. It varied with the make of lacquer, the size of the polishing bevels, and the temperature of the cutter. When this effect was added to the twenty-three degrees of the Westrex cutterheads, or the fifteen degrees of the proposed standard, major redesigns were needed in three dimensions. The Westrex 3D, the Neumann SX68, and the Ortofon cutterheads were the result. It proved possible to modify the Teldec by chamfering off a bottom edge and tilting it (to oversimplify greatly); thus all discs mastered after late 1964 should be to the fifteen-degree standard. But we should be prepared to mess about when playing stereo discs mastered before 1964. The best technique is to mount the pickup rigidly in its headshell, and arrange for the turntable to swing in gimbal mountings beneath it while listening to the vertical component of the sound, choosing an angle for minimum distortion.

A new type of cutting stylus was introduced in 1964 called the "Cappscoop." (Ref. 16). This was specifically intended to make the lacquer-springback more consistent, giving straighter groove-walls; but I have no experience of it or its results.

4.11 "Elliptical" and other styli

I have said several times that difficulties were caused when styli failed to trace the smaller wiggles in grooves, but I have not yet formally mentioned the solution. The smaller the zone of contact, the more accurately fine wiggles can be traced; but with small spherical tips the result is generally an increase in hiss and an increase in disc wear, because the contact takes place over a smaller area. Both problems can be ameliorated if we use a "bi-radial" stylus - that is, with a small size in one dimension and a large size in another. Edison's "button-shaped" stylus was one solution, and a sharp-tipped fibre was another. The so-called "elliptical" stylus was a third.

This is really only a trivial modification of Edison's idea. Edison made an oblate spheroid sit across the groove; the "elliptical stylus" comprises a conical tip rounded, not to a spherical shape, but to an ellipsoidal shape. When either Edison's or an ellipsoidal stylus sits in a V-shaped groove, the effect is the same. If you draw a horizontal cross-section of the stylus at the level of the points of contact, both styli are shaped like an ellipse; hence the shorter but inaccurate term, "elliptical stylus." Historically, the Ferranti ribbon pickup of 1948 was the first to be marketed with an elliptical sapphire stylus (Ref. 17), followed by a version of the Decca "ffrr" moving-iron pickup. Decca had been cutting full frequency-range coarsegroove shellac discs for four years, but towards the centre of the record the high frequencies were so crammed together that spherical styli could not separate them. An elliptical stylus not only extracted them better, but did so with less distortion.

In the case of Decca's stylus, the "dimensions" were 2.5 thou by 1.0 thou. By convention, this means that if you look at the horizontal cross-section at the point where the jewel sits across a V-shaped groove with walls of slope 45 degrees, the ellipse has a major axis of 5 thou (twice 2.5 thou) across the groove, and 2 thou (twice 1.0 thou) along the groove. It is also understood that the third dimension, up and down in the groove, also has a radius of 2.5 thou, but this axis may be slightly off vertical. When we speak of the "dimensions of an elliptical stylus," this is what we mean.

This turns out to be a useful compromise between a small spherical tip and a large spherical tip. In the example given, the stylus will follow the groove wiggles as satisfactorily as a 1 thou tip, while riding further up the groove walls (so there is less risk of the stylus running along the curved bottom of the groove, or hitting any noisy debris in the bottom). The disadvantage is that, although the area of contact is much the same, the ability to trace smaller wiggles can mean greater accelerations imparted to the stylus, and greater risk of disc wear on loud passages. Your organisation may like to consider the policy of using spherical tips for everyday playback purposes, and more sophisticated shapes for critical copying.

Reduced distortion can only be achieved if the major axis of the ellipse links two corresponding points of the opposite groove walls. It has been shown that the major axis has only to be in error by a few degrees for the reduction in distortion to be lost. Thus the pickup must be aligned for minimising both tracking and tracing distortions, particularly on inner grooves (section 4.4). The conventional alignment procedure assumes that the edges of the cutter which cut the two groove walls were oriented along the disc's radius. This was nearly always the case on discs mastered on cellulose nitrate, or the swarf wouldn't "throw" properly; but it is not unknown for wax-mastered discs to be substantially in error. A feather-edged cutter would cut a clean groove at almost any

angle. It may be necessary to "misalign" the arm, or the cartridge in the headshell, to neutralise the recorded tracing distortion.

At the time elliptical tips became popular, hi-fi enthusiasts were encouraged to spend time aligning their record-playing decks for optimum performance. This generally meant balancing record-wear against quality, and if you didn't want to damage any records in your collection, you needed to investigate the issues very thoroughly. But if you do not play vinyl or nitrate very much, most of the risk of wear can be circumvented by playing and transferring at half-speed.

Elliptical styli did not become commonplace until the mid-1960s. In the meantime, an attempt was made to reduce tracing distortion by pre-distorting the recorded groove. The RCA "Dynagroove" system was designed to neutralise the tracing distortion which occurred when a 0.7 thou spherical tip was used. (Ref. 18). So presumably that's what we should use for playing "Dynagroove" records today. But the "Dynagroove" system was also combined with a dynamic equalizer supposed to compensate for the Fletcher-Munson curves (a psychoacoustic phenomenon). The rationale behind this was essentially faulty, but the characteristics were publicly defined, and can be reversed. (Ref. 19)

If an elliptical tip does not have its vertical axis at the same angle as the vertical tracking angle, a phenomenon known as "vertical tracing distortion" occurs. This doesn't occur with spherical tips. I suspect the simultaneous existence of "vertical tracking" and "vertical tracing" distortion was responsible for the confusion between the words, but the terminology I have used is pedantically correct. Vertical tracing distortion *can* occur with mono lateral-cut discs, under extreme conditions of high frequencies and inner diameters. To put it in words, if the minor axis of the elliptical tip is tilted so that it cannot quite fit into the shorter modulations of the groove, results similar to conventional tracing distortion will occur. John R. T. Davies had some special styli made to play cellulose nitrate discs. These suffered from lacquer-springback even when the recording was made with a mono cutterhead, and for some reason the surface noise is improved by this technique as well. But a turntable in gimbals seems equally effective so long as the clearance beneath the cartridge is sufficient.

I said earlier that Edison's oblate spheroid was equivalent to an elliptical in a V-shaped groove; but it's not quite the same in a U-shaped groove. An elliptical will have the same difficulties fitting a U-shaped groove as a spherical, because looking in the direction along the groove, it appears spherical. The problem was solved by the "Truncated Elliptical" tip, a modern development only made by specialist manufacturers. It's an "elliptical" shape with the tip rounded off, or truncated, so it will always be driven by the groove *walls* and never the bottom. This shape is preferred for the majority of lateral coarsegroove records. (It even gives acceptable, although not perfect, results on most hill-and-dale records).

Although a range of sizes is offered, it is usually only necessary to change to avoid damage on a particular part of a groove wall, or to play lateral U-shaped grooves which have such a large radius that even a truncated tip is too small. Truncation can reduce the contact area and increase disc wear. Fortunately it is hardly ever needed for vinyl or cellulose nitrate records, which nearly always have V-shaped grooves.

We now reintroduce the lessons of "soft" styli, which had a large area of contact giving less hiss from particles in the pressed disc. Electronic synchronisation techniques permit us to play grooves with several styli of different sizes and combine the results. Thus, given a family of truncated ellipticals of different sizes, we emulate fibre needles without their disadvantages. I shall say more about this in section 0.

In the microgroove domain, the success of the elliptical stylus stimulated more developments, which are known collectively as "line-contact" styli. There were several shapes with different names. The first was the "Shibata" stylus, introduced in 1972 for playing the ultrasonic carriers of CD-4 quadraphonic discs (Section 10.16). The idea was to pursue lower noise, better frequency response, or lower wear, (or all three), by making contact with more of the groove walls.

But all line-contact styli suffer the same disadvantage. If the line of contact is not *exactly* correct - parallel to the face of the cutting stylus in the horizontal plane and fifteen degrees in the vertical - tracing distortion becomes very obvious. When everything is right they work well; but when anything is slightly misaligned, the result is disappointing. In 1980 an article in *Hi-Fi News* listed some of the types of line-contact stylus, mentioning that fundamentally faulty manufacturing principles and bad finish were adding to the difficulties. The author advocated the new "Van den Hul" stylus as being the solution; but a review of the very first such cartridge in the very same issue revealed that it had more distortion than half-a-dozen others. That review seems to have killed the idea for widespread use.

The trouble is that variations in the lacquer-springback effect and the tracking distortions of pivoted pickup arms made the ideal impossible to achieve without much fiddling. Cartridges with line-contact styli were expensive and delicate, and hi-fi buffs preferred fixed headshells, so fiddling was not made easier. So perfect reproduction was hardly ever achieved. It is significant that professionals have never used them. From the archival point of view, there is little need; most master tapes of the period still exist, and the subject matter is often available on compact digital disc. But clearly there is an avenue for exploration here. The reproduction of some older full-range records might well be improved, so for a general article on line-contact styli I refer you to Reference 20.

4.12 Other considerations

The above history should enable you to choose suitable styli and playing-conditions for yourself, so I do not propose to ram the points home by saying it all again. Instead, I conclude with a few random observations on things which have been found to improve the power-bandwidth product.

Many coarsegroove discs with V-shaped grooves have bottom radii which are smaller than the stylus sizes laid down by the British Standards Institution. Try a drastically smaller tip-radius if you can, but learn the sound of a bottoming stylus and avoid this. Not only does a small radius minimise pinch-effect and tracing distortions, but the bottom of the groove often survives free from wear-and-tear. This particularly applies to cellulose nitrates and late 78s with high recorded volumes. Indeed, these is some evidence that record companies did not change the sapphire cutter between a microgroove master and a 78 master. Eventually you will train your eye to tell the bottom radius of a groove, which will cut down the trial-and-error.

On the other hand, it sometimes happens that an outsized stylus is better. This is less common, because (all other things being equal) you will get increased tracing distortion, and there will be greater vulnerability to noise from surface scratches. But just occasionally the wear further down in the groove sounds worse. For various reasons, it seems unlikely we shall ever be able to counteract the effects of wear, so evasive action is advised. You can then concentrate upon reducing the distortion with elliptical or line-contact styli, and the noise with an electronic process.

Although my next point is not capable of universal application, there is much to be said for playing records with downward pressures greater than the pickup manufacturer recommends. To reduce record-wear, an audio buff would set his playing-weight with a test disc such as the Shure "Audio Obstacle Course", carrying loud sounds which might cause loss of groove contact. He would set his pickup to the minimum playing-weight to keep his stylus in contact with the groove walls at the sort of volumes he expected (different for classical music and disco singles!), thereby getting optimum balance between distortion and wear. But nowadays, little wear is caused by higher playing weights; most is caused when the grooves vibrate the stylus, not by the downward pressure. There can be several advantages in increasing the downward pressure for an archival transfer. The fundamental resonant frequency of the cantilever is increased (according to a one-sixth power law - Ref. 21), thereby improving the high frequency response. Clicks and pops are dulled, partly because the stylus can push more dirt aside, and partly because the cantilever is less free to resonate. But most important of all, the stylus is forced *into* the surface of the disc, thereby increasing the contact area and reducing the basic hiss. Obviously the operator must not risk causing irreparable damage to a disc; but if he is sufficiently familiar with his equipment, he will soon learn how far to go whilst staying within the elastic limits of the medium.

Shellac discs seem practically indestructible at any playing-weight with modern stereo pickup cartridges. Modern pickup arms are not designed for high pressures, but a suitably-sliced section of pencil-eraser placed on top of the head-shell increases the down-force with no risk of hangover. Pressures of six to ten grams often give improved results with such discs; special *low*-compliance styli should be used if they are available. With ultra-large styli, like those for Pathé hill-and-dale discs, it may even be necessary to jam bits of pencil-eraser between cantilever and cartridge to decrease the compliance further; twenty to thirty grams may be needed to minimise the basic hiss here, because the area of contact is so large.

Records should, of course, be cleaned before playback whenever practicable (see Appendix 1). But there are sometimes advantages in playing a record while it is wet, particularly with vinyl discs. Water neutralises any electrostatic charges, of course; but the main advantages come with discs which have acquired "urban grime" in the form of essence-of-cigarette-smoke, condensed smog, and sweaty fingerprints. Also, if previous owners have tried certain types of anti-static spray or other cleaning agents relying upon unconventional chemicals, there may be a considerable deposit on the groove walls which causes characteristic low-volume sounds. Conventional cleaning does not always remove these, because the sludge gets deposited back in the grooves before the record can be dried. Unfortunately it is impossible to give a rule here, because sometimes cleaning makes matters worse (particularly with nitrates - it may be essential to transfer each disc twice, once dry and once wet, and compare the results of the two transfers).

Centrifugal force often makes it difficult to play 78rpm discs wet. But for slower-speed discs, distilled water may be spread over the surface while it plays, perhaps with a minute amount of photographic wetting agent. The liquid can be applied through the outer casing of a ballpoint pen with the works extracted; this can be used as a pipette to apply the liquid, and as a ruler to spread it. Some types of disc have a "vinyl roar" which is caused when the stylus runs across the surface and excites mechanical resonances within the plastic. Although a proper turntable-mat and centre-clamp should eliminate the effect on most records, the liquid also helps. However, some transfer engineers have reported that dry playing of discs previously played wet can reveal a subsequent *increase* in surface noise. The author accepts no responsibility for damage to record or pickup!

I deliberately concentrated upon laterally-modulated records from section 4.7 onwards, but I shall now deal with a specific problem for hill-and-dale records. It is essential to take vertical tracking and vertical tracing into account of course, and strike a compromise between tracing distortion (caused by a large area of contact) and hiss (caused by a small area of contact). Even so, much even-harmonic distortion may remain, and in many cases this will be found to be recorded in the groove. The reason for this will be dealt with in section 4.15, where we look at electronic techniques for improving the power-bandwidth product.

Finally, the archivist should be aware that the metal intermediate stages in the disc-record manufacturing process – "master", "mother" and "stamper" - sometimes survive. Since these do not contain abrasives, the power-bandwidth product is usually better. I have no experience in playing metalwork myself, but a consensus emerged when I was researching this manual, which was that most people preferred to play fresh vinyl pressings rather than metal. There are a number of difficulties with metal - it is usually warped and lacks a proper-sized central hole, the nickel upsets the magnetic circuit of the pickup, you can have only "one bite of the cherry" whereas you may have several vinyl pressings, etc. However, as vinyl pressing plants are decommissioned, it will become increasingly difficult to get fresh vinyl pressings made, and the risk when a unique negative is clamped in a press by an inexperienced worker will increase. Until sound archives set up small pressing-plants, I think we are more likely to be playing metalwork in the future.

Pressing factories usually had the wherewithal to play a metal negative (with ridges instead of grooves), if only to be able to locate clicks or noise. The turntable must rotate backwards (see section 4.13), and the stylus must obviously have a notch so it can sit astride the ridge. Top quality isn't essential for factory-work; it is only necessary to locate problems without having to examine every inch of ridge under a microscope. The Stanton company makes a suitable stylus for their cartridges. In effect, it comprises two ordinary diamond tips side-by-side on the same cantilever. I am not aware that there are any options over the dimensions, so this could conceivably give disappointing results; but I must say the few I've heard sounded no worse than vinyl, and often better.

4.13 Playing records backwards

I shall now continue with a couple of "hybrid" topics. They combine mechanical techniques with electronic techniques. After that, the remaining sections will deal with purely electronic signal-processing. It has often been suggested that playing a record backwards and then reversing the transfer has some advantages. Among those cited are:

1. The opposite side of any steep wavefront is played, so wear has less effect.
2. Resonances and other effects which smear the signal in time are neutralised.
3. It is easier to extract the first milliseconds of modulation if the cutter has been lowered with sound on it.
4. It is easier to distinguish between clicks and music for electronic treatment.
5. If you are using fibre needles, the problems which would be caused by the needle being most-worn at the middle of the disc are ameliorated.
6. Needle-digs and other sources of repeating or jumping grooves are more easily dealt with.

Unfortunately the author simply does not agree with the first two reasons, although he has tried the idea several times. Worn records still sound worn (if the needle is tracing the groove correctly, of course). The theory of neutralising resonances is wrong. Even if electronic anti-resonance circuitry is proposed, the original waveform can only be recreated if the sound passes through the anti-resonant circuit *forwards*.

However, the other four arguments for playing a record backwards do have slightly more validity, but not much. In the case of argument (3), the writer finds that (on coarsegroove records, anyway) it is quicker to lower the pickup onto the correct place, repeating the exercise until it's done correctly! For argument (4), analogue click detectors work more efficiently because the circuitry is less confused by naturally-occurring transients, such as the starts of piano notes. But since all current analogue click detectors remove the click without replacing the original sound, they are not suited to archival uses. Computer-based declicking systems do not care whether the record is playing backwards or not; in effect, they shuttle the sound to and fro in RAM anyway. The writer has no experience of argument (5), because there is not yet a satisfactory electrical pickup using fibre needles, so you cannot reverse an electronic transfer anyway.

This leaves only the groove-jumping argument. For some records the reverse process can be very helpful. It will, of course, be necessary to use a reverse-running turntable, with a pivoted arm with a negative offset angle or a parallel-tracking system. Seth Winner, of the Rogers and Hammerstein Archives of Recorded Sound, has a conventional headshell with the cartridge facing backwards. He made this for playing disc-stamper negatives rather than records liable to groove-jumping. If his cartridge were to be used for groove-jumping, one would have to risk the cantilever being bent, because it will be compressed when it was designed to work under tension.

Also there are distinct disadvantages to the reverse-playing process. To start with, we need another turntable, or one which can be modified. A practical difficulty is that if the operator cannot understand the music, he may well miss other faults, such as wow, or lack of radius compensation (section 4.19). When some defects of equipment (such as tone-arm resonances) are reproduced backwards, the result is particularly distracting, because backward resonances cannot occur in nature.

To get the recording the right way round again, an analogue tape copy has to be reversed. For stereo, the left and right have to be swapped when the tape is recorded, so they will come out correctly on replay. Although I'd count it a luxury, if you were thinking of buying a digital audio editor, I'd advise getting one with the additional feature of being able to play a digital recording backwards while you were at it.

Since I haven't said much about groove-jumping, I shall now devote a paragraph to the subject, although I hesitate because any operator worth his salt should be able to invent ways round the difficulties much more quickly than I can advise him. The obvious way, adjusting the bias on the pickup-arm, causes the whole disc to be affected; so ideally you need a short-term aid. My method (which can also be applied to a parallel-tracking arm) is to apply some side-pressure through a small camel-hair paintbrush. With grossly-damaged records this isn't enough, so you may simply have to grab the cartridge lifting-handle between finger and thumb and *push*. This latter idea works best when you are copying at half-speed, which is the topic of the next section. You can't always get a transfer of archival quality under these conditions; so you may have to use your digital editor for its intended purpose, editing the results! For some notes on playing *broken* records, please see Appendix 1.

I shall now share an idea which I have not tried personally. We have seen that tracing distortions occur because a cutting-stylus does not have the same shape as a

replay stylus. Obviously, if we *play* a groove with a cutting stylus, we shall cut into it. But this wouldn't happen with a cutting stylus running backwards, and this could eliminate many kinds of tracing distortion. Extremely accurate matching between the shape and dimensions of the two styli would be needed, plus considerable reduction in the effective mass of the replay one to avoid groove deformation.

4.14 Half-speed copying

This is a technique which is useful for badly-warped or broken records which would otherwise throw the pickup out of the groove. It is particularly valuable for cylinders. It is almost impossible to get warped cylinders back to the original shape, and most of them rotate faster than discs anyway. The solution is to transfer the item at half the correct speed to a system running at half the desired sampling frequency.

The principal disadvantage is that the *low*-end responses of all the equipment have to be accurate beyond their normal designed limits. Another is that the natural momentum of all moving parts is lower, so speed variations in the copying equipment are always higher. It is true that, given good modern equipment, the errors are likely to be swamped by those of the original media; but you should remember the danger exists.

4.15 Distortion correction

You will note that this is the first significant area in which the word "maybe" occurs. I shall be talking about processes which have yet to be invented. I don't intend to infuriate you, but rather to show where techniques are possible rather than impossible. In the archival world time should not be of the essence, so you could leave "possible but not yet practical" work until a later date.

At present, harmonic and intermodulation distortion are faults which never seem to be reverse-engineered electronically. In principle, some types of such distortion could easily be undone; it seems the necessary motivation, and therefore the research, hasn't happened. I can only recall one piece of equipment which attempted the feat during playback - the Yamaha TC800 cassette-recorder of 1976. It certainly made the Dolby tone (Section 8.4) sound better; but personally I could hear no difference to the music!

In the circumstances, I can only advise readers to make sure as little distortion as possible comes off the medium at source, because (as we shall see later) there are electronic ways of dealing with noise. Until someone breaks the mould, we must assume that retrospective distortion-removal will never be possible, and therefore we must concentrate upon it at source.

Harmonic distortion is, in practice, always accompanied by intermodulation distortion. For a reasonably complete survey of this idea I refer you to Reference 22; but in the meantime I will explain it briefly in words. If *two* frequencies are present at the same time, say m and n, we not only get harmonics ($2m$, $3m$, $4m$ . . . and $2n$, $3n$, $4n$ . . . , the conventional "harmonic distortion"), but we also get "sum-and-difference frequencies" ($m+n$, $m-n$, $2m-n$, $2n-m$, etc). The latter case is called "intermodulation distortion." Subjectively, the worst case is usually ($m-n$), because this means extra frequencies appear which are lower in pitch than the original sounds, and very conspicuous. They are often called "blasting." *If* they have come from this source (they could also come from transient effects in the power-supply of the recording amplifier), the

only hope for removing them without filtering is to generate equal-and-opposite sum-and-difference frequencies by reverse-engineering the original situation.

Gazing into my crystal ball, I can see no reason why distortion-removal should always remain impossible. One can visualise a computer-program which could look at a musical signal in one octave, pick up the harmonic and intermodulation products in other octaves, and by trial-and-error synthesise a transfer-characteristic to minimise these. By working through all the frequency-bands and other subsequent sections of sound, it should be possible to refine the transfer characteristic to minimise the overall distortions at different volumes and frequencies. It would be an objective process, because there would be only one transfer characteristic which would reduce all the distortion products in the recording to a minimum, and this would not affect naturally-occurring harmonics. If future research then finds a transfer characteristic which is consistent for several separate recordings done with similar equipment, we might then apply it to an "objective copy." I admit that, unless there is a paradigm shift because of a completely new principle, it would mean billions of computation-intensive trials. But computer-power is doubling roughly each year, so ultimate success seems inevitable.

The conventional approach - reverse-engineering the original situation - would depend upon having access to the sound with the correct amplitudes and relative phases. I have already mentioned the importance of phase in section 2.13. When we come to frequency equalisation in later chapters, I shall be insisting on pedantically correct ways of doing equalisation for this reason.

The first progress is likely to be made in the area of even-harmonic distortion, which occurs on recorded media which do not have a "push-pull" action. These include hill-and-dale grooves, the individual channels of a stereo groove, and unilateral optical media. Sometimes these show horrendous distortion which cries out for attention. Sometimes they are essentially reproduction problems, but at other times the recording medium will cause varying load on a cutter, meaning distortion is actually recorded into the groove.

In the late 1950s even harmonic tracing distortion was heard (for the first time in many years) from stereo LP grooves. The two individual groove walls did not work together to give a "push-pull" action to a stylus; they acted independently, giving only a "push" action. It was suggested that record manufacturers should copy a master-nitrate with the phases reversed so as to exactly neutralise the tracing distortion when the second reproduction took place. Fortunately, this was not necessary; as we saw in section 4.11, new types of playback styli were developed to circumvent the difficulty. And there was very little *recorded* distortion, because by that time the cutterheads were being controlled by motional negative feedback, which virtually eliminated distortion due to the load of the nitrate.

Many decades before, some manufacturers of hill-and-dale records did actually copy their masters, incidentally cancelling much of this sort of distortion. Pathé, for instance, recorded on master-cylinders and dubbed them to hill-and-dale discs (Ref. 23), and at least some of Edison's products worked the opposite way, with discs being dubbed to cylinders. And, of course, "pantographed" cylinders were in effect dubbed with phase-reversal. So there are comparatively few cases where hill-and-dale records have gross even-harmonic distortion. It is only likely to occur with original wax cylinders, or moulded cylinders made directly from such a wax master.

The fact that it was possible to "correct" the even harmonic distortions shows that it should be easy with electronics today; but archivists must be certain such processes do not corrupt the odd harmonics, and this means we need more experience first.

The CEDAR Noise reduction System includes an option which reduces distortion. This uses a computerised "music model" to distinguish between music and other noises. Details have not yet been made public, so it is impossible to assess how objective the process is, so I cannot yet recommend it for archive copies.

4.16 Radius compensation

Edison doggedly kept to the cylinder format long after everyone else, for a very good engineering reason. With a disc rotating at a constant speed, the inner grooves run under the stylus more slowly than the outer grooves, and there is less room for the higher frequencies. Thus, all things being equal, the quality will be worse at the inner grooves. Cylinders do not have this inconsistency. Earlier we saw some of the difficulties, and some of the solutions, for *playing* disc records. But I shall now be dealing with the *recording* side of the problem, and how we might compensate it.

A "feather-edged" cutter was not affected by the groove speed. Such cutters were used for wax recording until the mid-1940s. With spherical or "soft" styli, there would be problems in *reproduction*; but today we merely use a bi-radial or line-contact stylus to restore the undistorted waveform. We do not need to compensate for the lack of high frequencies electrically.

The problem only occurred when the cutter did not have a sharp edge, e.g. because it had a polishing bevel. Here the medium resisted the motion of the cutter in a manner directly proportional to its hardness. For geometrical reasons it was also inversely proportional to the groove speed, and inversely proportional to the mechanical impedance of the moving parts. (A stiff armature/cutter will be less affected than a floppy one. A cutter with motional feedback has a high mechanical impedance). Finally, the effect was also dependent upon the size of the polishing bevel and the temperature of the wax or lacquer at the point of contact. All these factors affected the high-frequency response which was cut into the disc.

Thus, even with perfect groove contact, we may notice a high-frequency loss today. The effect will be worst on a recording "cut cold" in lacquer, using a duralumin-and-sapphire coarsegroove cutting-tool in a wide-range cutterhead with low mechanical impedance. In practice, the effect seems worst on semi-pro nitrate 78s recorded live in the 1950s.

Because of the complexity of the problem, and because no systematic analysis was done at the time, the effect cannot be reversed objectively. On any one record, it's usually proportional to the groove speed; but human ears work "logarithmically" (in octaves rather than wavelength). The subjective effect is usually imperceptible at the outside edge of the disc. It is often inaudible half-way through; but the nearer the middle, the worse it starts to sound.

We do not know precisely when recording engineers started compensating for the effect as they cut the master-disc. It is thought Western Electric's Type 628 radius-compensator circuit was in use by 1939. Before this date, the official upper-frequency limits of electrical recording systems prevented the effect from demanding much attention. After 1939, it can be assumed that commercial master-disc cutting incorporated radius compensation in some form. We may have to play the pressings with line-contact or elliptical styli to minimise the pinch-effect distortion, but this should not affect the *intended* frequency-range; compensation for the *recorded* losses will have been performed by the mastering engineer.

For other discs, the present-day transfer operator should compare the inner and outer radii. The usual procedure is to assume that the outside edge suffers *no* radius loss, and compensate for the high-frequencies at other radii *by ear* on the *service-copy only*. The operator will certainly have to do this if the subject matter requires the sides to be seamlessly joined!

Because the effect is wavelength-dependent, the compensation circuit should ideally vary the *frequency* of the slope continuously, not the slope itself. There is a limit to the compensation possible without making drastic increases in hiss and harmonic distortion. When we consider this, we observe that objective compensation is impossible for another reason. The transfer operator must use subjective judgement to balance the effects and minimise them for the listener.

The author knows of only one organisation which treated radius-compensation scientifically, and unfortunately its research was based on a different foundation. During the second world war, the BBC was attempting to stretch the performance of its nitrate lacquer disc-cutting operation to 10kHz, and the engineers considered the *whole* system (recording *and* reproduction) together. So far as reproduction was concerned, they settled on a standard stylus (2.5 thou spherical sapphire) and a standard pickup (the EMI Type 12 modified so its fundamental resonance was 10kHz), and they devised radius-compensation which gave minimum distortion when nitrates were played with this equipment. And higher frequencies were ignored, because the landline distribution system and the characteristics of double-sideband amplitude-modulation transmission usually eliminated frequencies above 10kHz anyway. The compensation was designed for cold cutting of V-shaped grooves into cellulose nitrate blanks. The result was a family of resonant circuits in the recording electronics, each with a different resonant frequency and peak level. An electrical stud system (like a stud fader) switched between these circuits about five times during every inch of recorded radius. (Ref. 24). This continued until the BBC abandoned coarsegroove nitrate discs in about 1965.

From today's viewpoint, this puts us in a dilemma. It would seem that we should play such discs with a 2.5 thou spherical sapphire in an EMI Type 12 cartridge; but this is a destructive instrument by today's standards, and it will damage the disc. Furthermore the BBC assumed the nitrate had consistent hardness and elasticity. Several decades later the material has altered considerably, so accurate reconstruction of the intended situation is impossible anyway. Finally it may be impossible for academics running a sound-archive to recover the original *intended* sound, because of the tradeoffs made to minimise side-changes after the sound was broadcast with limited bandwidth.

The current policy at the British Library Sound Archive is to compensate only for the steady-state recording characteristics (which we shall be considering in chapter 5). We generally play the discs with the usual truncated elliptical styli to recover the maximum power-bandwidth product, but we do not attempt to neutralise the resonant artefacts at high frequencies, which are audible (but not severe) under these conditions. It is possible that some form of adaptive filtering may analyse the high-frequency spectrum and compensate it in future; in the meantime we have preserved the power-bandwidth product, which is the fundamental limitation.

The remainder of this chapter is concerned with the state-of-the-art in audio restoration technology, but can only be considered to be so at the time of writing. While much of the information will inevitably become outdated, it may still remain instructive and of some future use.

BOX 4.17
METHODS OF REPLACING CLICKS

1. The first is to cut both the click and the sound which was drowned by it, and to pull the wanted sounds on either side together in time. This is how tape editors have worked for the past forty years. Not only does it destroy part of the original waveform, but in extreme cases it can destroy tempo as well.
2. Another answer is to replace the click with *nothing*. Certainly, it is true that leaving a hole in the music is less perceptible than leaving the click; but we can hardly call it "restoring the original sound" - at least, if we mean the objective sound-wave rather than the sensation.
3. Another answer is to synthesise something to fill the gap. A very popular method is the "two-band method" (where there are two processors, one dealing with high frequencies, which leaves little holes as before, and one dealing with low frequencies, which holds the instantaneous voltage throughout the gap). This is subjectively less noticeable, but again you cannot call it "restoring the original sound."
4. John R. T. Davis was the inventor of the "Decerealization" technique, which emulates this process. It involves a quarter-inch analogue magnetic tape of the clicky disc. A special jig which incorporates a tape-head and an extremely accurate marking-device holds the tape. Its dimensions are such as to permit a "stick-slip action" as the tape is pulled by hand. The operator listens on a selected loudspeaker, and as the individual short segments of sound are reproduced, the click stands out conspicuously. After the position is marked, the surface layer of the tape is scraped off where the click is. Although very labour-intensive, this remains the best way to deal with some types of material, because the operator can scrape off different degrees of oxide, thus creating the effect of the previous method with variable crossover frequencies for each click. In addition, when you can't hear the click, the waveform isn't attacked.

5. Another technique is to take a piece of sound from somewhere else in the recording and patch it into the gap. This technique was first described by D. T. N. Williamson, and although automatic devices using the idea have been proposed, they have never appeared. (It was envisaged that sound delayed by a few milliseconds could be patched into place, but it was found difficult to change to the delay-line without a glitch). Manual equivalents of the principle have been used successfully by tape editors. It has the "moral" advantage that you can say nothing has been synthesised. All the sounds were made by the artist!
6. More elaborate synthesis is used by the digital computer-based noise reduction methods "No-Noise" and "CEDAR." They analyse the sound either side of the click, and synthesise a sound of the same spectral content to bridge the gap.
7. The final solution to the question "what do we replace the click with" only works if you have two copies of a sound recording and each of them suffers from clicks in different places. Then we can take the "best of both" without interpolating the waveform.

4.17 Electronic click reduction

The elimination of clicks has been a tantalising goal for more than half a century, because it is a relatively simple matter to *detect* a click with simple electronic techniques. The problem has always been: What do we replace the click with? (see Box 4.17)

All but the last method disqualify themselves because they do not pretend to restore the original sound waves; but if you need more details, please see Ref. 25. Unfortunately, there are relatively few pieces of equipment which can reduce noise without affecting the wanted waveform. In fact there are so few that I must mention actual trade-names in order to make my points; but I should remind readers these pieces of apparatus will be displaced in time. Some of them may be needed only occasionally, and may be hired instead of purchased; or recordings might be taken to a bureau service to cut down the capital expenses. You will need to know the various options when formulating your strategy.

The most important objective technique is Idea (7) in the Box, which is employed as the "first stage" of the Packburn Noise Reduction System (Packburn Electronics Inc, USA; Ref. 26). This is an analogue processor widely used in sound archives, and it has three stages. The first is used when a mono disc is being played with a stereo pickup, and the machine chooses the quieter of the two groove walls. It cannot therefore be used on stereo records.

Analysis of the actual circuit shows that it only attenuates the noisier groove wall by 16dB, so the description I have just given is something of an oversimplification; but it is certainly effective. The result is a little difficult to quantify, because it varies with the nature of the disc-noise and how one measures the result; but on an unweighted BBC Peak Programme Meter an average EMI shellac pressing of the inter-war years will be improved by about ten decibels. And, as I say, the waveform of the wanted sound is not, in principle, altered.

I should, however, like to make a few points about the practical use of the circuit. The first is that if we play one groove-wall instead of both groove walls, we find ourselves with a "unilateral" medium. Thus we risk even-harmonic distortion, as we saw in section 4.15. Actually, there is a mid-way position on the Packburn such that the two groove walls are paralleled and the whole thing functions as a lateral "push-pull" reproduction process. Theoretical analysis also shows that optimum noise reduction occurs when the groove walls are paralleled whenever they are within 3dB of each other. The problem is to quantify this situation.

The manufacturers recommend you to set the "RATE" control so the indicator-lights illuminate to show left groove-wall, right groove-wall, and lateral, about equally. I agree; but my experience with truncated-elliptical styli is that there is very little even-harmonic distortion reproduced from each groove wall anyway. You shouldn't worry about this argument; there are other, more-significant, factors.

The next point is that, in the original unmodified Packburn, control-signals broke through into the audio under conditions of high gain, giving a muffled but definite increase in background noise which has been described subjectively using the words "less clarity" and "fluffing". Therefore the RATE control must be set to the maximum which actually improves the power-bandwidth product and no more. My personal methodology is based on playing HMV Frequency Test Disc DB4037, which we shall be considering in chapter 5. Using a high frequency test-tone, we can easily hear the best noise reduction happens when the three light-emitting diodes are lit for about the same overall time. Thus the manufacturer's recommendation is confirmed. Do this on real music, and the optimum

power-bandwidth is assured, even though it is less easy to hear the side-effects. Now that "The Mousetrap" manufactured in the UK by Ted Kendall has replaced "The Packburn," this problem has been eliminated by the use of high-speed insulated-gate field effect transistors (IGFETs).

Another point is that, if the machine is to switch between the two groove walls successfully, the wanted sound on those two groove walls must be identical in volume and phase. (Otherwise the switching action will distort the waveform). The Packburn therefore has a control marked "SWITCHER - CHANNEL BALANCE." When you are playing a lateral mono disc, you switch the main function switch to VERTICAL, and adjust this control to *cancel* the wanted signal. Then, when you switch back to LATERAL, the two groove walls will be going through the processor at equal volumes.

All this is made clear in the instruction-book. But what if you *cannot* get a null? In my view, if the wanted sound is always audible above the scratch, there's something wrong which needs investigating. Assuming it isn't a stereo or fake-stereo disc, and you can get a proper cancellation on known mono records (which eliminates your pickup), then either the tracking angle is wrong (most of Blumlein's discs, section 6.31 below), or you've found a record made with a faulty cutterhead (e. g. Edison-Bell's - section 6.16 below).

The former fault can be neutralised by slewing the pickup cartridge in its headshell. The latter faults have no cures with our present state of knowledge, but cures may be discovered soon, which would be important because sometimes there is useful power-bandwidth product in the vertical plane. In the meantime, all you can do is slew the cartridge as before in an attempt to cancel as much sound as possible, and then try the Packburn in its usual configuration to assess whether its side-effects outweigh the advantage of lower surface-noise.

To decide between the two groove walls, the machine needs access to undistorted peak signals at frequencies between 12kHz and 20kHz. It has been said that even the best stereo analogue tape copy of a disc will mar the efficiency of the unit, because it "clips" the peaks or corrupts their phase-linearity, and it is rather difficult to keep azimuths (section 9.6) dead right. This makes it difficult to treat a record unless you have it immediately beside the Packburn. Actually, I do not agree; I have even got useful noise reduction from a stereo *cassette* of a disc. But certainly the Packburn isn't at its best under these conditions.

But digital transfers seem "transparent" enough. So it is practicable to use a two-channel digital transfer for the "archive" (warts-and-all) copy, provided no disc de-emphasis is employed (sections 3.5 or 6.23). Meanwhile, for objective and service copies it is best to place the Packburn following a flat pickup preamplifier with the usual precautions against high-frequency losses. Any frequency de-emphasis must be placed *after* the Packburn. (This is indeed how the manufacturers recommend the unit should be used).

The "second stage" of the Packburn is "the blanker", a device for removing clicks which remained after the first stage, either because both groove walls were damaged at the same place, or because it was a stereo disc. The Packburn's blanker rapidly switches to a low-pass filter, whose characteristics are designed to minimise the subjective side-effects (as paragraph (3) of Box 4.17. It does not restore the original sound wave, so it should only be used for service copies. Likewise, the "third stage" comprises a quite good non-reciprocal hiss reduction system (chapter 10), but this too alters the recorded waveform, so it too should be confined to service copies. To remove the remaining hiss and crackle

whilst keeping the waveform, we must use alternative techniques; but the Packburn "first stage" is a very good start.

There are two such alternative techniques. One is to emulate the action of the Packburn first stage, but using two different copies of the same record. I shall be talking about this idea here and in section 4.20. The other is to use computer-based digital processing techniques to synthesise the missing sound.

The first idea is still in the development stage as I write, but the principle of its operation is very simple. Two copies of a disc pressed from the same matrix are played in synchronism. If the discs are mono, each goes through a "Packburn first-stage" (or equivalent). The difficult part is achieving and keeping the synchronism, for which the geometrical errors must be kept very low; but once this is achieved, a third Packburn first-stage (or equivalent) cleans up the result. Using the same example as I had in section 4.16, the result is a further 8dB improvement in signal-to-noise ratio. The noise coming from each disc is not actually steady hiss (although it usually sounds like it), but a very "spiky" hiss which responds to the selection process. If it had been pure white-noise, equal on the two copies but uncorrelated, the improvement would only be 3dB. (Which would still be worth having). Isolated clicks are generally completely eliminated, and no synthesis of the waveform is involved.

For this principle to work, the two discs have to be synchronised with great accuracy - better than 0.2 milliseconds at the very least - and this accuracy must be maintained throughout a complete disc side. Although digital speed-adjustment techniques exist, we saw in section 3.4 these have disadvantages which we should avoid if we can. So use a deck with minimal geometrical errors. For example, use a parallel-tracking arm whose pickup is pivoted in the plane of the disc, or provided with an effective means of keeping it a constant distance above the disc surface, so warps do not have an influence. The sampling-frequency of the analogue-to-digital converter is then "locked" to the turntable speed; there are other reasons in favour of doing this, which I shall mention towards the end of section 5.5. In the British Library Sound Archive's case, a photoelectric device has been used to look at the stroboscope markings at the edge of the turntable, giving a 100Hz output. The result is frequency-multiplied by 441, giving a 44.1kHz clock-signal for the analogue-to-digital converters. The transfers of the two discs are done using the same stylus and equalisation, and at the same level, through a Packburn first-stage. The results are transferred to a digital audio editor and adjusted until they play in synchronism. The result is fed back through another Packburn at present, although a digital equivalent is being written to avoid unnecessary D-A and A-D conversions.

It has been found advantageous to combine the two discs using methods which operate on different frequency-bands independently. The Packburn only switches in response to noises in the range 12 – 20kHz. But if we have uncorrelated low frequency noises (e.g. rumble introduced during pressing), the switching action will generate sidebands, heard as additional clicks. In a prototype digital equivalent of the Packburn First Stage, we divide the frequency range into discrete octaves and treat each octave separately. The switching action takes place in each octave at the optimum speed for minimising sideband generation, and of course we get the quieter of the two grooves at *all* frequencies (not just 12-20kHz). We also get the version with the least distortion-effects in each band. The wanted waveform is never touched; all that happens is that background-noises and distortions due to the reproduction process are reduced. But at least two "originals" must be available.

We return now to when we have only one "original." It is always possible to combine two plays of the *same* disc with different-sized styli, using the technology I have just described. This imitates the action of a "soft" stylus!

Several systems synthesise the missing section of waveform (previously drowned by the click) and insert it into place. Most digital processes use a technique known as the Fast Fourier Transform, or FFT, to analyse the wanted sound either side of the click. This is a speedy algorithm for a binary computer; in audio work, it is some hundreds of times faster than the next best way of doing the job, so it can run usefully even on a desktop microcomputer. (Ref. 27). When the click is eliminated, random digits are re-shaped according to the FFT analysis, and when the inverse FFT is performed, the missing waveform is synthesised. Both "No-Noise" and "CEDAR" offer realtime implementation, so the operator can switch between the processed and unprocessed versions and check that nothing nasty is happening to the music. Both replace the clicks with synthesised sound, so in principle we are not actually restoring the original waveform; but it's a matter of degree. Experiments can be set up taking known waveforms, adding an artificial click, and seeing what sort of job the computer does of synthesising the original. The result may be judged aurally, or visually (on a waveform display of some sort).

The CEDAR people have various pieces of hardware and software for click removal. This includes a computer-based platform offering several powerful restoration algorithms (not just clicks), free-standing boxes which cannot be used for any other purpose (the cheaper one has no analogue-to-digital or digital-to-analogue converters), and cards for slotting into a SADiE PC-based hard-disk editor. The last-mentioned is usually the most recent version, since it is easier to make both "beta versions" and proven hardware.

Although "real-time," the free-standing DC.1 unit may require the signal to be played through the machine three times, since three different algorithms are offered; they should be performed starting with the loudest clicks and ending with the quietest. CEDAR very bravely admit that the DC.1 process does not always synthesise the waveshape correctly for a long scratch, but in 1994 they claimed it was correct for clicks up to fifty samples long. CEDAR have been extending the number of samples; the latest version is in the range 250-300 samples. (This clearly shows archivists must log the version-number of the software). If the results were put to a scientific test on both aural and visual grounds with 100% successful results, there would presumably be no objection to using the algorithm for objective copies as well as service copies.

Unfortunately, since one must start with the biggest clicks, and the DC.1 sometimes blurs these (making it more difficult for future processes to detect them), there are relatively few records for which the DC.1 gives archivally-acceptable results. Malcolm Hobson's solution is to run his process several times from hard disc in batch mode (this avoids having to accumulate several R-DATs which must work in real time). He starts with an FFT looking for high frequency transients less than six samples long (these are almost bound to be components of crackle), then interpolates these (which is certain to give faithful results for such small clicks). The process then works *upwards* towards larger clicks. Each time the surrounding music has less crackle, so interpolation is easier. However, much loving care-and-attention is needed for the benign replacement of the largest clicks, which may be done manually. So the trade-off is between leaving blurred clicks and possibly-inaccurate interpolation.

The No-Noise process is obliged to work in conjunction with a Sonic Solutions editing-system, which could be a restriction for some customers; but it is possible to concatenate several processes (for example, de-clicking, de-hissing, *and* equalisation), and

run them *all* in real-time. This helps the operator to listen out for unwanted side-effects to any one of the processes. No-Noise has an option to mark the start and end of long clicks manually, and then do a trial synthesis of the missing signal, which you can adopt if you are satisfied. I have heard it do a convincing job on many thousands of missing samples, but I do not know how "accurate" this was. Although it has nothing to do with click removal, this process seems to be the best way to synthesise sound for a sector of a broken record which has vanished. This will probably never be an objective technique; but over the years many similar jobs have been done in the analogue domain. No-Noise can help in *two* ways, firstly by synthesising the missing sound, and secondly by performing edits non-destructively.

CEDAR's computer-platform system and their free-standing de-crackle unit Type CR.1 offer another process. To oversimplify somewhat, the recording is split digitally into two files, one through a "music model" and the other comprising everything else. It is then possible to listen to the "music model" on its own, and adjust a control so that even the biggest clicks are eliminated. (This file lacks high frequencies and has various digital artefacts along with the music, but it is easy to listen for loud clicks if they are there). When a satisfactory setting has been achieved which eliminates the loudest clicks but goes no further, the two files are recombined. This process has been found empirically to reduce many of the effects of harmonic distortion, as I mentioned in section 4.15.

As we go to press, audio engineers are exploring other mathematical strategies for synthesising missing data. So far, these studies seem to comprise "thought experiments", with no "before-and-after" comparisons being reported. The only one to have appeared is the newly developed SASS System (invented by Dr. Rudolf Bisping). Prony's Method is used to analyse the music and express it as a sum of exponentially-decaying frequencies, which enables complete remodelling of the amplitude spectrum, including signals which change in pitch and notes which start or stop during the click. To do all this requires a computer some hundreds of times more powerful than hitherto. The SASS System has a dedicated architecture involving many transputers; but once again I have not had an opportunity to test it for "accuracy."

Interpolation for replacing the biggest clicks is still not reliable. But it is well-known that interpolation is easier on sounds which change slowly, and that clicks appear subjectively louder on these same sounds. I consider we need a click-replacement process which automatically adapts itself to the subject matter. To end with a crudely-expressed dream, we need an interpolation strategy which automatically knows the difference between clicks during slow organ music, and clicks during a recording of castanets.

4.18 Electronic hiss reduction

No-Noise, CEDAR, and SASS also offer "hiss-reduction" algorithms. I wish to spend some time talking about these, because they offer conspicuously powerful methods of reducing any remaining noise; but frankly I am sure they are wrong for archival storage purposes.

The idea is to carve up the frequency-range into a number of bands, then reduce the energy in each band when it reaches a level corresponding that of the basic hiss. The process can reduce hiss very spectacularly; but it can cut audible low-level signals fainter than the hiss, so listeners sometimes complain there is "no air around" the performance. At its best it can reduce so much noise that it is possible to extract wanted high-frequency sounds which are otherwise inaudible, thereby apparently making a dent in the power-bandwidth principle (section 2.2).

I must also report that an analogue equivalent was being marketed by Nagra at one point (the Elison Model YSMA 18 with eighteen frequency bands); but it did not become available for some reason, which was a pity as it could be operated more intuitively than any digital equivalents.

Unfortunately these processes must make use of psychoacoustics to conceal side-effects. The width of the sample being analysed (in both the time and frequency domains), the amplitude below which attenuation can take place, the degree of attenuation, the times of response and recovery, and the volume at which reproduction is assumed to take place may all need to be taken into account. To make matters worse, recent psychoacoustic experiments suggest that our ears work differently when listening to speech as opposed to music. Most hiss-reduction units have user-adjustable controls for some of these factors. Although offered on a try-it-and-see basis, this subjective approach rules it out for archival applications. The controversies about records which have been processed by both No-Noise and CEDAR are usually attributable to the fact that the operator and the consumer have different psychoacoustic responses and/or listening conditions. Nevertheless, psychoacoustic measurements have made strides in the last decade, and many of the factors can now be quantified with precision, and, equally importantly, with a clear understanding of the variances.

The Fast Fourier Transform requires the number of frequency bands to be an exact power of two, with linear spacing. At present, No-Noise uses 2048 bands and CEDAR uses 1024, giving bands about 11 and 22 Hertz wide respectively when the sampling-frequency is 44.1kHz. This is not the optimum from the psychoacoustic point of view; it is well known that the human ear deals with frequency bands in quite a different way.

To reduce hiss whilst (apparently) leaving the wanted sounds intact, wanted sounds must "mask" unwanted ones. The unit of masking is the "bark". Listening tests suggest that the human ear has a linear distribution of barks at frequencies below about 800Hz, and logarithmic above that. Thus any computer emulating the masking properties of the human ear needs a rather complex digital filter. Furthermore, the slopes of the filtered sections of the frequency range are asymmetrical, and vary with absolute volume. The last-mentioned parameter has been circumvented by Werner Deutsch et. al. (Ref. 28), whose team chose values erring on the side of never affecting the wanted sound ("overmasking"). In my view, this algorithm is the best available method of reducing hiss while leaving the (perceived) wanted sound untouched. It is surprisingly powerful. Even when the hiss and the music are equal in volume, the hiss can be reduced by some 30dB; but its quality then becomes very strange. There is also the difficulty that, owing to the vital contribution of psychoacoustics, any frequency or volume changes must be made *before* the process, not after it.

Bisping divides the spectrum into 24 bands that correspond to critical bands in Corti's organ of the inner ear, and hiss-removal is inhibited when it is isn't needed. The trouble is that many experts in psychoacoustics consider 24 bands an oversimplification!

Such processes might be applied to the service copies of recordings meant to be heard by human adults. (But *not* to other sounds, or to sounds which might go through a second process involving audio masking). Even so, the correct archival practice must surely be to store the recording complete with hiss, and remove the hiss whenever it is played back.

We may yet find that the human ear has evolved to make the best use of the information presented to it, with little room for manoeuvre. We already know that the threshold of hearing is almost exactly at the level where the Brownian-movement of individual air molecules lies. Thus we might find that our pitch-detection and our

tolerance to background-noise have evolved together to give a performance which cannot be improved. If so, we could never reduce wideband hiss to reveal completely inaudible sounds. But I very much look forward to further developments, because they might permit a real breakthrough in sound restoration. The limitations of the power-bandwidth principle could be breached for the very first time.

4.19 Eliminating rumble

Apart from the use of linear filters (which affect the wanted sound, of course), there have been very few attempts to neutralise the low-pitched noises caused by mechanical problems in the cutting machine. Not all such noises are capable of being removed "objectively," but there are a few exceptions. Usually these occur when the machine was driven by gears (as opposed to idler-wheels, belts, or electronic servo-systems). Here the pattern of rumble may form a precise relationship with the rotational speed of the cylinder or disc. Using the principle of digital sampling being locked to rotational speed, as mentioned in section 0 above, it is possible simply to add together the sounds from each turn of the record. When this is done, you may often find a consistent pattern of low-frequency rumble builds up, which may be low-pass filtered and then subtracted from each of the turns in the digital domain to reduce the noises without affecting the wanted sound. This is particularly valuable when you're trying to get some bass back into acoustic recordings (Chapter 11).

4.20 De-thumping

This section deals with the side-effects of large clicks when played with many practical pickup arms. The click may shock-excite the arm or disc into resonances of its own, so that even when the click is eliminated, a low-frequency "thud" remains. From one or two cases I have known, I suspect the actual cartridge may also be excited into ringing-noises at much higher frequencies (a few kiloHertz). Sometimes these exhibit extremely high Q-factor resonances, which cause long "pinging" noises. In the ordinary course of events these artefacts are not audible, because they are masked by the click itself. Only when the click is removed does the artefact become audible.

Quite frankly, the best solution is not to generate the thumps in the first place. Very careful microscopic alignment of cracked records may be needed to ensure that the pickup is not deviated from the centre-line of its travel. The cartridge may need to have silicone grease packed in or around it to reduce its tendency to make mechanical or electrical signals of its own. The pickup arm must have well-damped mechanical resonances; alternatively, a "parallel-tracking" arm may be used (section 4.2). This suspends the cartridge above the record in a relatively small and light mechanism, and, all things being equal, has less tendency to resonate. (Some parallel-trackers are capable of dealing with misaligned cracks which would throw a pivoted tone-arm, because they are guided by a relatively inert motor mechanism. This is probably the best way to play a warped and/or broken disc which cannot be handled any other way).

The Tuddenham Processor not only removes clicks; it also has a de-thump option, which applies a transient bass-cut with an adjustable exponential recovery. However, this is a subjective process, which should only be applied to the service-copy.

CEDAR have an algorithm for de-thumping which relies on having a large number of similar thumps from a cracked record. When twenty or thirty are averaged, the components of the wanted sound are greatly reduced, leaving a "template" which can be subtracted from each one. It does not significantly corrupt the original waveform, so it has its place. Sounds intended for such treatment should have the basic clicks left *in place*, as CEDAR uses them to locate the thumps. The makers of the SADiE digital editor can supply a de-thump card for the machine, but I simply do not have experience of its principles or its ease of operation.

4.21 Future developments

As I write, the industry is excited by the possibilities of adaptive noise-cancellation. This is similar to the hiss-reduction processes I mentioned in section 4.18, except that instead of using a fixed sample of hiss to define the difference between hiss and music, it can be a dynamic process. (Ref. 29). Given a sample of "pure" noise varying with time, the computer can (in theory) do a Fast Fourier Transform of the noise, and use it to subtract the appropriate amount of energy from the signal.

The makers envisage it could be used on the following lines. If it were desired to eliminate the background noise of speech picked up in an aircraft cockpit, for example, the usual noisy recording would be made, plus another track of pure aircraft noise (from another microphone). In theory, it would then be possible to sample the pure aircraft noise and use it to reduce the noise behind the speech, without having to rely upon phase-accurate information. This is actually an old idea (Ref. 30), which hitherto has been used by the US military for improving the intelligibility of radio-telephone communication. Only now is enough computing power becoming available for high-fidelity applications. For once, sound archivists don't have to plead their special case. It is an application with many uses in radio, films, and television, and I anticipate developments will be rapid.

With mono disc records there are new possibilities, because "pure noise" is largely available. It is possible to extract a noise signal from a lateral disc by taking the vertical output of the pickup, although the rumble of the recording-turntable usually differs between the two planes. It offers the only hope for ameliorating cyclic swishing noises when there is only one copy of the disc, or when all surviving copies are equally affected. Some of the effects of wear might also be neutralised, although I wouldn't expect to get all the sound back; indeed, we might be left with conspicuous low-frequency intermodulation distortion. Certain types of modulation noise on tape recordings could also be reduced, by splitting a tape track in two and antiphasing them to provide a "clean noise" track. Since it is possible to insert timeshifts into either signal, tape "print-through" might even be reduced. In chapter 11 I shall be considering the feasibility of dynamic expansion; this would almost certainly have to be done in conjunction with adaptive noise-cancellation to conceal the effect of background noise going up and down. But it seems these applications must always be subjective processes, only to be considered when drastic treatment is essential for service copies.

I should stress that adaptive noise-cancellation still has not achieved success in audio. One attempt to reduce long disc clicks failed, because there was insufficient processing power to analyse rapidly-varying noise. Disc clicks are distinguished from wanted sound *because* they change so rapidly.

At one time CEDAR were developing an algorithm which emulated the dual-processing method described in section 0 above, although it did not actually take the

wanted sound from two copies. No "hard-lock" synchronisation was involved, so it could be used on wild-running transfers from two widely different sources. The reason it is not yet available is that it was very computation-intensive, and did not always offer satisfactory results because of difficulties synchronising recordings in the presence of noise. In the case of disc records, the two copies each underwent click-reduction first, so the whole point of dual processing (to avoid interpolation) was missed. (Refs. 31 and 32). Nevertheless, this process might be used for reducing the basic hiss of already-quiet media, such as two magnetic tapes of the same signal. But it will always be computation-intensive. One can only hope that increased processing power and further research might make this process successful.

4.22 Recommendations and conclusion

This ends my description of how to recover the power-bandwidth product from grooved media, but I have not yet formally stated my views about what the three versions should comprise.

The archive copy should be a representation of the groove reproduced to constant-velocity characteristics (see chapter 6) using a stereo pickup, so that the two groove walls are kept separate. Ideally, there should be several separate transfers done with styli of different sizes, to provide samples at different heights up the groove wall. It is natural to ask how many transfers this should be. Experiments with the earliest version of the program mentioned in section 0 have been done, in which additional counting procedures were inserted to quantify the number of samples taken from each transfer. This was checked by both digital and analogue methods of measuring the resulting signal-to-noise ratio. All three methods suggest that, for coarsegroove shellac discs played with truncated elliptical styli, four such transfers should be done with styli whose larger radii differ by 0.5 thou. Softer media, such as vinyl or nitrate, will have a greater commonality between the transfers, because the styli will penetrate deeper into the groove walls; so four is the most which will normally be needed. Of course, this means a lot of data must be stored; but if you accept that Program J1* does the job of combining the transfers optimally, you can use this, and still call it an "archive copy." To help any future anti-distortion processes, the stylus dimensions, and the outer and inner radii of the disc grooves, should be logged. And I remind you that absolute phase must be preserved (section 2.11).

For the objective copy, the same procedure should be followed, except that known playing-speeds (chapter 5 and recording characteristics (chapter 6) should be incorporated. Clicks may be eliminated, so long as accurate interpolation of the previously-drowned waveform occurs. It is natural to ask what tolerance is acceptable. My answer would be to do some before-and-after tests on the declicker; if subtracting "after" from "before" results in no audible side-effects, then the waveforms were synthesised accurately enough. But I recognise readers might have other ideas.

For the service copy, radius compensation may be applied, speed adjustments for artistic reasons can be incorporated, hiss-reduction may be considered, and sides may be joined up where necessary (section 13.2).

I hope my peek into the future won't leave you cross and frustrated. Digital techniques are admittedly costly and operationally cumbersome, but there has been enormous progress in the last few years. By the time you read these words, the above paragraphs are certain to be out-of-date; but I include them so you may see the various

* Editors' note: program J1 was probably written by Peter Copeland but has not been found.

possibilities. Then you can make some sensible plans, start the work which can be done now, and put aside the jobs which may have to wait a decade or two.

REFERENCES

1: Franz Lechleitner, "A Newly Constructed Cylinder Replay Machine for 2-inch Diameter Cylinders" (*paper*), Third Joint Technical Symposium "Archiving The Audio-Visual Heritage," Ottawa, Canada, 5th May 1990.

2: Percy Wilson, "Modern Gramophones and Electrical Reproducers," (*book*), (London: Cassell & Co., 1929), pp. 126-128.

3: P. J. Packman, British patent 23644 of 1909.

4: The earliest reference I have found is an anonymous article in Talking Machine News, Vol. VII No. 89 (May 1909), page 77.

5: Carlos E. R. de A. Moura, "Practical Aspects of Hot Stylus," Journal of the Audio Engineering Society, April 1957 Vol. 5 No. 2, pp. 90-93.

6: A. M. Pollock, Letter to the Editor. London: *Wireless World*, April 1951, page 145.

7: S. Kelly, "Further Notes on Thorn Needles." *Wireless World*, June 1952, pages 243-244.

8: A. M. Pollock, "Thorn Gramophone Needles." *Wireless World*, December 1950, page 452.

9: H. E. Roys, "Determining the Tracking Capabilities of a Pickup" (*article*), New York: Audio Engineering Vol. 34 No. 5 (May 1950), pp. 11-12 and 38-40.

10: S. Kelly, "Intermodulation Distortion in Gramophone Pickups." *Wireless World*, July 1951, pages 256-259.

11: F. V. Hunt, "On Stylus Wear and Surface Noise in Phonograph Playback Systems." Journal of the Audio Engineering Society, Vol. 3 No. 1, January 1955.

12: J. A. Pierce and F. V. Hunt, J.S.M.P.E, Vol. 31, August 1938.

13: J. Walton, "Stylus Mass and Distortion." Paper presented to the Audio Engineering Society Convention in America in October 1962, but only published in printed form in Britain. *Wireless World* Vol. 69 No. 4 (April 1963), pp. 171-178.

14: Roger Maude: "Arnold Sugden, stereo pioneer." London: *Hi-Fi News*, October 1981 pages 59 and 61. (Includes complete discography)

15: John Crabbe: "Pickup Problems, Part Two - Tracking Error," *Hi-Fi News*, January 1963, pp. 541-545.

16: Richard Marcucci, "Design and Use of Recording Styli," J.A.E.S., April 1965 pp. 297-301.

17: *Wireless World*, April 1948 page 135.

18: John Crabbe: "Dynagroove Hullabaloo," *Hi-Fi News*, November 1963 pages 417 and 419, and December 1963 pages 521 and 523.

19: Harry F. Olsen: "The RCA Victor Dynagroove System" (*paper*), Journal of the Audio Engineering Society, April 1964, pp. 203-219.

20: Basil Lane, "Improving groove contact," *Hi-Fi News*, August 1980 pages 75-77.

21: C. R. Bastiaans, "Factors affecting the Stylus/Groove Relationship in Phonograph Playback Systems," Journal of the Audio Engineering Society, (1967?), pages 107-117.

22: "Cathode Ray": "More Distortion . . . What Causes Musical Unpleasantness?" (*article*), *Wireless World* Vol. 61 No. 5 (May 1955), pp. 239-243.

23: Girard and Barnes, "Vertically Cut Cylinders and Discs" (*book*), pub. The British Library Sound Archive.
24: For a general article on the BBC philosophy, see J. W. Godfrey and S. W. Amos, "Sound Recording and Reproduction" (*book*), London: Iliffe & Sons Ltd (1952), page 50 and pages 80-82. Details of the effect of the circuit for coarsegroove 33s and 78s on the BBC Type D disc-cutter may be found in BBC Technical Instruction R1 (October 1949), page 8; and there is a simplified circuit in Fig. 1.
25: Adrian Tuddenham and Peter Copeland, "Record Processing for Improved Sound" (*series of articles*), "Part Three: Noise Reduction Methods," London, *Hillandale News* (the journal of the City of London Phonograph and Gramophone Society), August 1988, pages 89 to 97.
26: Richard C. Burns, "The Packburn Audio Noise Suppressor" (*article*), Sheffield, The Historic Record No. 7 pages 27-29. (March 1988).
27: The Fast Fourier Transform was invented in several forms by several workers at several times, and there does not seem to be a definitive and seminal article on the subject. For readers with a maths A-level and some programming experience with microcomputers, I recommend Chapter 12 of the following:
William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling: "Numerical Recipes - The Art of Scientific Computing," (*book*), Cambridge University Press (1989). This is available in three editions, the recipes being given in three different computer languages.
28: Werner A. Deutsch, Gerhard Eckel, and Anton Noll: "The Perception of Audio Signals Reduced by Overmasking to the Most Prominent Spectral Amplitudes (Peaks)" (preprint), AES Convention, Vienna, 1992 March 24-27.
29: Francis Rumsey, "Adaptive Digital Filtering" (*article*), London: Studio Sound, Vol. 33 No. 5, pp. 34-5. (May 1991).
30: (Pioneer adaptive noise cancellation paper)
31: Saeed V. Vaseghi and Peter J. W. Rayner, "A New Application of Adaptive Filters for restoration of Archived Gramophone Recordings" (paper), I.E.E.E Transcriptions on Acoustics Speech and Signal Processing, 1988 pages 2548-2551.
32: UK Patent Application GB 2218307.

5 Speed setting

5.1 Introduction

Few things give more trouble than setting the speed of an anomalous recording correctly. There are many factors in the equation, and often they are contradictory. This writer therefore feels it is important, not only to take corrective action, but to document the reasons why a decision has been made. Without such documentation, users of the transferred recording will be tempted to take further corrective action themselves, which may or may not be justified - no-one knows everything!

I must (with respect) point out that "psychoacoustics" can often play a dominant role in speed-setting. Personally, I can't do the following trick myself, but many musicians consistently and repeatedly get a sensation that something is "right" when they hear music at the correct pitch. They usually can't articulate how they know, and since I don't know the sensation myself, I can't comment; but it's my duty to point out the potential traps of this situation. It's a craft that musicians will have *learnt*. I am not saying that such musicians are necessarily "wrong". I am, however, saying that musical pitch has changed over the years, that actual performances will have become modified for perfectly good scientific reasons, and yet hardly anybody has researched these matters. Ideally therefore, analogue sound restoration operators should make themselves aware of *all* the issues, and be prepared to make judgements when trying to reach "the original sound" or "the *intended* original sound."

When we come to make an objective copy, there are two types of analogue media which need somewhat different philosophies. One occurs when the medium gives no indication of where a particular sound is located, the main examples being full-track magnetic tape and magnetic wire. In these cases it is impossible even to add such information retrospectively without sacrificing some of the power-bandwidth product, because there are no sprockets, no pulses, no timecode, nor any "spare space" to add them. But other cases have a "location-mechanism by default." For example, we could refer to a particular feature being at "the 234th turn of the disc record". It is very possible that future digital processes may use information like this; and ideally we should not sacrifice such information as we convert the sound to digital. During this chapter we shall see that it is often impossible to set a playing-speed with greater accuracy than one percent. In which cases, it may be advantageous to invoke a "digital gearbox" to lock the rotational speed of the disc with the sampling-frequency of the digital transfer, so the rotations of the disc do not evaporate from the digital copy.

Pure sound operators are sometimes unaware that a very close lock is vital in some circumstances, so I shall define that word "lock." It means that the speed of the original medium and the speed of the transferred sound must match to a very tight tolerance (typically one part in a million). This is much tighter than most ordinary sound media can do; so we may need to create our own "digital gearbox," especially for digital signal-processes downstream of us. And this means we may have to do some creative thinking to establish a suitable gear-ratio.

On the other hand, it is impossible to "lock" analogue media which gradually change in speed with a fixed "gearbox." But obviously some form of gearbox is essential for a sound medium intended to accompany moving pictures, since it is always implied

that "Pictures are King," and sound must follow in synchronism, even if it's actually the wrong speed! As an illustration of the point I am trying to make, to provide consistently reproducible sound for a film running at 24 frames per second, we could multiply the frame-rate by 2000 (making 48000), and clock our analogue-to-digital converter from that.

5.2 History of speed control

I shall start with a brief history of speed-control in the history of analogue sound recording, and ask you to bear it in mind as different situations come up.

The very earliest cylinder and disc machines were hand-cranked, but this was soon found unsatisfactory, except for demonstrating how pitch varied with speed !

Motor-drive was essential for anything better. Early recorders of the 1880s and 1890s were powered by unregulated DC electric motors from primitive electrochemical cells. Several percent of slow speed drift is the normal result.

Clockwork motors, both spring and weight powered, quickly replaced electric motors because of the attention which such early batteries demanded. But mainsprings were less reliable than falling weights, so they tended to be used only where portability was essential (location recording), and for amateur use. The centrifugal governor was adopted at the same time to regulate such motors; the one in the surviving acoustic lathe at the EMI Archive, which is weight-powered, is made to exactly the same pattern as in spring gramophones. Oddly enough, a spring would have given better results for edge-start disc-records. According to Hooke's law, the tension in a spring is proportional to its displacement, so there was more torque at the start of the recording, precisely where it was most needed. Yet professional disc recordists actually preferred the weight system, justifying the choice with the words "Nothing is more consistent than gravity."

The governor could be adjusted within quite wide limits (of the order of plus or minus twenty percent). Most commercial disc records were between 70 and 90rpm, with this range narrowing as time progressed. Likewise, although location or amateur cylinders might well differ widely from the contemporary standards (section 5.4), they were often quite constant within the recording itself.

In the late 1920s alternating-current electric mains became common in British urban areas, and from the early 1930s AC electric motors began to be used for both disc and film recording. These motors were affected by the *frequency* of the supply. During cold winters and most of the second World War, frequencies could vary.

BBC Radio got into such trouble with its programmes not running to time that it adopted a procedure for combating it, which I shall mention in detail because it provides one of the few *objective* ways of setting speeds that I know. It applies to "Recorded Programmes" (i.e. tape and disc recordings with a "R. P. Number", as opposed to BBC Archive or Transcription recordings) made within a mile or two of Broadcasting House in London. (I shall be mentioning these further in section 6.52). The various studios took line-up tone from a centrally-placed, very stable, 1kHz tone-generator (which was also used for the "six pips" of the Greenwich Time Signal). When a recording was started, a passage of this line-up tone was recorded, not only to establish the programme volume (its main purpose), but as a reference in case the frequency of the supply was wrong. When the disc or tape was played back, it was compared with the tone at the time, and the speed could be adjusted by ear with great accuracy. We can use this technique today. If you are playing a "BBC Recorded Programme" and you have an accurate 1kHz tone-

source or a frequency-counter, you can make the recording run at precisely the correct speed. This applies to recordings made at Broadcasting House, 200 Oxford Street, and Bush House; you can recognise these because the last two letters of the R. P. Reference Number prefixes are LO, OX and BU respectively. But *do not* use the system on other BBC recordings, made for example in the regions. The master-oscillators at these places were deliberately made different, so that when engineers were establishing the landlines for an inter-regional session, they could tell who was who from the pitch of the line-up tone. But there was an internal procedure which stated that *either* accurate 1kHz tone was used, *or* tone had to be at least five percent different. So if you find a line-up tone outside the range 950Hz - 1050Hz, ignore it for speed-correction purposes.

To continue our history of speed-stability. Transportable electrical recording machinery became available from the late 1930s which could be used away from a mains supply. It falls into three types. First we have the old DC electric motor system, whose speed was usually adjusted by a rheostat. (For example, the BBC Type C disc equipment, which a specialist can recognise from the appearance of the discs it cut. In this case an electronic circuit provided a stroboscopic indicator, although the actual speed-control was done manually by the engineer). Next we have mains equipment running from a "transverter" or "chopper", a device which converted DC from accumulators into mains-voltage A.C. (For example, the machine used by wildlife recordist Ludwig Koch. These devices offered greater stability, but only as long as the voltage held up). Finally we have low-voltage DC motors controlled by rapidly-acting contacts from a governor. (For example, the EMI "L2" portable tape recorder). All these systems had one thing in common. When they worked, they worked well; but when they failed, the result was catastrophic. The usual cause was a drop in the battery voltage, making the machine run at a crawl. Often no-one would notice this at the time. So you should be prepared to do drastic, and unfortunately empirical, speed correction in these cases.

It wasn't until the "transistor age" that electronic ways of controlling the speed of a motor without consuming too much power became available, and in 1960 the first "Nagra" portable recorder used the technology. From the late 1960s electronic speed control became reliable on domestic portable equipment. Similar technology was then applied to mains equipment, and from about 1972 onwards the majority of studio motors in Britain began to be independent of the mains frequency.

But do not assume your archive's equipment is correct without an independent check. I insist: an *independent* check. Do not rely on the equipment's own tachometers or internal crystals or any other such gizmology. I regret I have had too much experience of top-of-the-range hardware running at the wrong speed, even though the hardware itself actually insists it is correct! You should *always* check it with something else, even if it's only a stroboscopic indicator illuminated by the local mains supply, or a measured length of tape and a stopwatch. As an example of this problem, I shall mention the otherwise excellent Technics SL.1200 Turntable, used by broadcasters and professional disc-jockeys. This is driven from an internal crystal; but the same crystal generates the light by which the stroboscope is viewed. The arithmetic of making this work forces the stroboscope around the turntable have 183 bars, rather than the 180 normally needed for 50Hz lighting in Europe. So the actual speed may be in error, depending how you interpret the lighting conditions!

5.3 History of speed-control in visual media

I must also give you some information about the methods of speed-control for film and video. Pure sound archivists may run into difficulties here, because if you don't understand the source of the material, you may not realise you have something at the wrong speed. And if your archive collects picture media as well, you need a general idea of the history of synchronisation techniques, as these may also affect the speed of the soundtrack. But if your archive doesn't use any sound which has accompanied moving pictures in the past, I suggest you jump to Section 5.4.

As this isn't a history of film and video, I am obliged to dispense with the incunabula of the subject, and start with the media which the European archivist is most likely to encounter.

In silent-film days, cameras were generally hand-cranked, and the intended speed of projection was between twelve and twenty frames per second. For the projector, the shutter had two blades, or sometimes three; this chopped up the beam and raised the apparent frequency of flicker so that it was above the persistence of vision. Moving scenes did, of course, jerk past slower than that, but this was acceptable because the brightness of cinema screens was lower in those days, and the persistence of vision of human eyes increases in dim light.

But there was a sudden and quite definite switch to a higher frame-rate when sound came along. This happened even when the sound was being recorded on disc rather than film, so it seems that the traditional story of its being necessary to allow high audio frequencies onto the film is wrong. I suspect that increasing viewing-standards in cinemas meant the deficiencies of slower speeds were very obvious to all. So, with motorised cameras now being essential if the accompanying soundtrack were to be reproduced with steady pitch, the opportunity was taken for a radical change. Anyway, nearly all sound films were intended for projection at 24 frames per second, and all studio and location film crews achieved this speed by appropriate gearing in conjunction with A.C. motors fed from a suitable A.C. supply.

There were two basic methods which were used for both recording and projection; they ran in parallel for a couple of years, and films made by one process were sometimes copied to the other. They were optical sound (which always ran at the same speed as the picture, whether on a separate piece of celluloid or not), and coarsegroove disc (always exactly 33 1/3rpm when the picture was exactly 24 frames per second). Most film crews had separate cameras and sound recorders running off the same supply, and the clapper-board system was used so the editor could match the two recordings at the editing-bench.

Because location-filming often required generous supplies of artificial light, location crews took a "power truck" with them to generate the power; but this does not mean the A.C supply was more vulnerable to change, because of a little-known oddity. The speed of 24 frames per second had the property of giving steady exposure whether the camera looked at 50Hz or 60Hz lighting. If however the lights and the camera were running off separate supplies, there was likely to be a cyclic change in the film exposure, varying from perhaps once every few seconds to several times per second. Location electricians therefore spent a significant amount of time checking that all the lights plus the picture and sound cameras were all working off the same frequency of supply, wherever the power actually came from.

Since the invention of the video camera, this fact has been "rediscovered", because 50Hz cameras give "strobing" under 60Hz lights and vice-versa. So, since the earliest days of talkies, location power supplies have never been allowed to vary, or strobing would occur. Thus we can assume fairly confidently that feature-films for projection in the cinema are all intended to run at 24 frames per second. And whether the sound medium is sepmag, sepopt, commag, comopt, or disc, it can be assumed that 24 fps working was the norm - until television came along, anyway.

But when television did come along in the late 1940s, this perfection was ruined. Television was 25 fps in countries with 50Hz mains, and 30 fps in countries with 60Hz mains, to prevent "rolling hum-bars" appearing on the picture. This remains generally true to this day. (For the pedantic, modern colour NTSC pictures - as used in America and Japan - are now 29.97 frames per second). In Europe we are so used to 50Hz lighting and comparatively dim television screens that we do not notice the flicker; but visiting Americans often complain at our television pictures and fluorescent lighting, as they are not used to such low frequencies of flicker at home.

Before the successful invention of videotape (in America in 1956), the only way of recording television pictures was "telerecording" (US: "kinetoscoping") - essentially filming a television screen by means of a film camera. Telerecording is still carried out by some specialists, the technique isn't quite dead. All current television systems use "interlacing," in which the scene is scanned in two passes called "fields" during one frame, to cut down the effect of flicker. To record both halves of the frame equally, it is necessary for the film camera to be exactly "locked" to the television screen, so that there are precisely 25 exposures per second in 50Hz countries, and 30 (or later 29.97) exposures per second in 60Hz countries.

So whether the sound medium is comopt, commag or sepmag, the speed of a telerecording soundtrack is always either 25, 30 or 29.97 frames per second. Thus, before you can handle an actual telerecording, you must *know* that it is a telerecording and not a conventional film, and run it on the appropriate projector. A cinema-type projector will *always* give the wrong speed.

The real trouble occurs when film and video techniques are mixed, for example when a cinema film is shown on television. We must not only know whether we are talking about a film or a telerecording, but we must also know the country of transmission.

In Europe, feature films have always been broadcast at 25 frames per second. Audio and video transfers from broadcast equipment are therefore a little over four percent fast, just under a semitone. Thus music is always at the wrong pitch, and all voices are appreciably squeaky. There is quite a lot of this material around, and unless you know the provenance, you may mistakenly run it at the wrong speed. Keen collectors of film music sometimes had their tape-recorders modified to run four percent faster on record, or four percent slower on playback; so once again you have to know the provenance to be certain.

Meanwhile, cinema films broadcast in 60Hz countries are replayed at the right speed, using a technique known as "three-two pulldown." The first 24fps frame is scanned *three* times at 60Hz, taking one-twentieth of a second; the next frame is scanned *twice*, taking one-thirtieth of a second. Thus two frames take one-twelfth of a second, which is correct. But the pictures have a strange jerky motion which is very conspicuous to a European; but Americans apparently don't notice it because they've always had it.

Optical films shot specifically for television purposes usually differed from the telerecording norm in America. They were generally 24 frames per second like feature

films. This was so such films could be exported without the complications of picture standards-conversion. But in Europe, cameramen working for TV have generally had their cameras altered so they shoot at 25 frames per second, like telerecordings. Thus stuff shot on film for television is at the right speed in its country of origin; but when television films cross the Atlantic in either direction they end up being screened with a four percent error. Only within the last decade or so have some American television films been shot at 30 frames per second for internal purposes.

Up to this point I have been describing the conventional scenario. To fill in the picture, I'm afraid I must also mention a few ways in which this scenario is wrong, so you will be able to recognise the problems when they occur.

Although feature-films are made to world-wide standards, there was a problem when commercial videos became big business from about 1982 onwards. Some videos involving music have been made from American films (for example Elvis Presley movies), and these were sometimes transferred at 24 fps to get the pitch right. This was done by what is laughingly called "picture interpolation." To show twenty-four frames in the time taken for twenty-five, portions of the optical frame were duplicated at various intervals; this can be seen by slow-motion analysis of the picture. The sound therefore came out right, although the pictures were scrambled. In cases of doubt, still-frame analysis of a PAL or SECAM video can be used as evidence to prove the audio is running correctly!

More often, it is considered preferable *not* to distort the picture. Here I cannot give you a foolproof recipe. My present experience (1999) suggests that most of the time the sound is four percent fast; but I understand some production-houses have taken to using a "Lexicon" or "harmonizer" or other device which changes pitch independently of speed (Ref. 1). Thus if the musical or vocal pitch is right *and* there are no video artefacts, it *may* mean that the *tempo* of the performance is wrong.

But there have now been two more twists in the story. Sometimes American television material is shot on film at 24 fps, transferred to 30 fps videotape for editing and dubbing into foreign languages, and then subjected to electronic standards-conversion before being sent to Europe. This gives the right speed of sound, but movement-anomalies on the pictures; but again, you can regard the presence of movement anomalies as evidence that the sound is right. The second twist came with Snell and Wilcox's "DEFT" electronic standards-converter, which has sufficient solid-state memory to recognise when "three-two pulldown" has taken place. It is then possible to reverse-engineer the effect to "two-two pulldown," and copy steady images to a video recorder running at 48 fields per second, ready for transmission on a conventional video machine at 50Hz. Again, the steady pictures warn you something is wrong with the sound.

5.4 Setting the speed of old commercial sound records

In setting the speed of professional audio recordings, my opinion is that the first consideration (which must predominate in the absence of any other evidence) is the manufacturer's recommended speed. For the majority of *moulded* commercial cylinder records, this was usually 160 revolutions per minute; for most coarsegroove records, it was about 80rpm until the late 1920s and then 78rpm until microgroove came along; for magnetic tapes, it was usually submultiples of 60 inches per second. (Early "Magnetophon" tapes ran a little faster than 30 inches per second, and this is thought to apply to EMI's earliest master-tapes made before about 1951. Ref. 2).

Unfortunately it isn't always easy for the modern archivist to discover what the recommended speed actually was. It does not always appear on the record itself, and if it is mentioned at all it will be in sales literature, or instructions for playback equipment made by the same company.

The recommended speeds of Edison commercial pre-recorded cylinders have been researched by John C. Fesler (Ref. 3). The results may be summarised as follows:

1888-1892: 100rpm

Mid-1892 to at least 1st November 1899: 125rpm

June 1900 to the beginning of moulded cylinders: 144rpm

All moulded cylinders (late 1902 onwards): 160rpm.

It is also known that moulded cylinders by Columbia were intended to revolve at 160rpm, and this forms the "baseline" for all moulded cylinders; so do not depart from 160rpm unless there is good reason to do so.

The following is a list of so-called 78rpm discs which weren't anywhere near 78, all taken from "official" sources, contemporary catalogues and the like.

Berliner Gramophone Company. Instructions for the first hand-cranked gramophones recommended a playing-speed of about 100rpm for the five-inch records dating from 1890-1894, and 70rpm for the seven-inch ones dating from about 1894-1900. But these are only "ballpark" figures.

Brunswick-Cliftophone (UK) records prior to 1927 were all marked 80rpm. Since they were all re-pressings from American stampers, this would appear to fix the American Brunswicks of this time as well.

Columbia (including Phoenix, Regal, and Rena): according to official company memoranda, 80rpm for all recordings made prior to 1st September 1927, from both sides of the Atlantic; 78rpm thereafter. But I should like to expand on this. The company stated in subsequent catalogues that Columbia records should be played "at the speed recommended on the label." This is not quite true, because sometimes old recordings were reissued from the original matrixes, and the new versions were commonly labelled "Speed 78" by the printing department in blissful ignorance that they were old recordings. The best approach for British recordings is to use the matrix numbers. The first 78rpm ones were WA6100 (ten-inch) and WAX3036 (twelve-inch). At this point I should like to remind you that I am still talking about *official* speeds, which may be overridden by other evidence, as we shall see in sections 5.6 onwards. Note too that Parlophone records, many of which were pressed by Columbia, were officially 78.

Edison "Diamond discs" (hill-and-dale): 80rpm.

Grammavox: 77rpm. (The Grammavox catalogue was the pre-war foundation for the better-known UK "Imperial" label; Imperial records numbered below about 900 are in fact Grammavox recordings).

Vocalion: All products of the (British) Vocalion company, including "Broadcast", "Aco", and "Coliseum", and discs made by the company for Linguaphone and the National Gramophonic Society, were officially 80rpm.

Finally, there are a few anomalous discs with a specific speed printed on the label. This evidence should be adopted in the absence of any other considerations.

There also has to be a collection of "unofficial" speeds; that is to say, the results of experience which have shown when *not* to play 78s at 78.

It is known that for some years the US Victor company recorded its master-discs at 76rpm, so they would sound "more brilliant" when reproduced at the intended speed of 78rpm. (This seems to be a manifestation of the syndrome whereby musicians tune their instruments sharp for extra brilliance of tone). At the 1986 Conference of the Association of Recorded Sound Collections, George Brock-Nannestad presented a paper which confirmed this. He revealed the plan was mentioned in a letter from Victor to the European Gramophone Company dated 13th July 1910, when there had been an attempt to get agreement between the two companies; but the Gramophone Company evidently considered this search for artificial brilliance was wrong, and preferred to use the same speeds for recording and playback. George Brock-Nannestad said he had confirmed Victor's practice upon several occasions prior to the mid-1920s.

Edison-Bell (UK) discs (including "Velvet Face" and "Winner") tended to be recorded on the high side, particularly before 1927 or so; the average is about 84rpm.

Pathé recordings before 1925 were made on master-cylinders and transferred to disc or cylinder formats depending upon the demand. The speed depends on the date of the matrix or mould, not the original recording. The earliest commercial cylinders ranged from about 180rpm to as much as 200rpm, and then they slowed to 160 just as the company switched to discs in 1906. The first discs to be made from master-cylinders were about 90rpm, but this is not quite consistent; two disc copies of Caruso's famous master-cylinders acquired by Pathé, one pressed in France and one in Belgium, have slightly different speeds. And some Pathé disc sleeves state "from 90 to 100 revolutions per minute." But a general rule is that Pathé discs without a paper "label" (introduced about 1916) will have to run at about 90rpm, and those with a paper label at about 80. The latter include "Actuelle," British "Grafton," and some "Homochord."

In 1951 His Master's Voice issued their "Archive Series" of historic records (VA and VB prefixes). The company received vituperation from collectors and reviewers for printing "SPEED 78" in clear gold letters upon every label, despite the same records having been originally catalogued with the words "above 78" and "below 78."

Quite often official recommended speeds varied from one record to the next. I will therefore give you some information for such inconsistent commercial records.

Odeon, pre-1914. The English branch of the Odeon Record company, whose popular label was "Jumbo", were first to publicise the playing speeds for their disc records. They attempted to correct the previous misdemeanours of their recording-engineers in the trade magazine *Talking Machine News* (Vol.VI No.80, September 1908), in which the speeds of the then-current issues were tabulated.

Subsequently, Jumbo records often carried the speed on the label, in a slightly cryptic manner (e.g. "79R" meant 79 revolutions per minute), and this system spread to the parent-company's Odeon records before the first World War. We don't know nowadays precisely how these speeds were estimated. And, although I haven't conducted a formal survey, my impression is that when an Odeon record *didn't* carry a speed, it was often because it was horribly wrong, and the company didn't want to admit it.

Gramophone, pre-1912. The leading record company in Europe was the Gramophone Company, makers of HMV and Zonophone records. In about 1912 they decided to set a standard of 78rpm, this being the average of their contemporary catalogue, and they also conducted listening experiments on their back-catalogue. The resulting speed-estimates were published in catalogues and brochures for some years afterwards; for modern readers, many can be found in the David and Charles 1975 facsimile reprint "Gramophone Records of the First World War."

Experienced record-collectors soon became very suspicious of some of these recommendations. But if we ignore one or two obvious mistakes, and the slight errors which result from making voices recognisable rather than doing precise adjustments of pitch, the present writer has a theory which accounts for the most of the remaining results. Gramophones of 1912 were equipped with speed-regulators with a central "78" setting and unlabelled marks on either side. It seems there was a false assumption that one mark meant one revolution-per-minute. But the marks provided by the factory were arbitrary, and the assumption gave an error of about 60 percent; that is to say, one mark was about two-and-a-half rpm. So when the judges said "Speed 76" (differing from 78 by two), they should have said "Speed 73" (differing from 78 by five). If you're confused, imagine how the catalogue editors felt when the truth began to dawn. It's not surprising they decided to make the best of a bad job, and from 1928 onwards the rogue records were listed simply as "above 78" or "below 78". Nevertheless, it is a clear indication to us today that we must *do* something!

Speeds were usually consistent within a recording-session. So you should not make random speed-changes between records with consecutive matrix numbers unless there is good reason to do so. But there are some exceptions. Sometimes one may find alternative takes of the same song done on the same day with piano accompaniment and orchestral accompaniment; these may appear to be at different speeds. This could be because the piano was at a different pitch from the orchestra, or because a different recording-machine was used. When a long side was being attempted, engineers would sometimes tweak the governor of the recording-machine to make the wax run slower. I would recommend you to be suspicious of any disc records made before 1925 which are recorded right up to the label edge. These might have been cut slower to help fit the song onto the disc.

I must report that so-called 78rpm disc records were hardly ever recorded at exactly 78rpm anyway. The reason lies in the different mains frequencies on either side of the Atlantic, which means that speed-testing stroboscopes gave slightly different results when illuminated from the local mains supply, because the arithmetic resulted in decimals. In America a 92-bar stroboscope suggests a speed of 78.26087rpm; in Europe a 77-bar stroboscope suggests a speed of 77.922078rpm. The vast majority of disc recording lathes

then ran at these speeds, which were eventually fixed in national (but not *international*) standards. From now on, you should assume 78.26 for American recordings and 77.922 for European recordings whenever I use the phrase "78rpm disc." A similar problem occurs with 45rpm discs, but not 33 1/3s; stroboscopes for this speed can be made to give exact results on either side of the Atlantic.

5.5 Musical considerations

My policy as a sound engineer is to start with the "official" speed, taking into account the known exceptions given earlier. I change this only when there's good reason to do so. The first reason is usually that the pitch of the music is wrong.

It's my routine *always* to check the pitch if I can, even on a modern recording. (Originals are often replayed on the same machine, so a speed error will cancel out; thus an independent check can reveal an engineering problem). I only omit it when I am dealing with music for films or other situations when synchronisation is more important than pitch. Copies of the two "Dictionaries of Musical Themes" may be essential (unfortunately, there isn't an equivalent for popular music). Even so, there are a number of traps to setting the speed from the pitch of the music, which can only be skirted with specialist knowledge.

The first is that music may be transposed into other keys. Here we must think our way into the minds of the people making the recording. It isn't easy to transpose; in fact it can only be done by planning beforehand with orchestral or band accompaniments, and it's usually impossible with more than two or three singers. So transposition can usually be ruled out, except for established or VIP soloists accompanied by an orchestra; for example, Vera Lynn, whose deep voice was always forcing Decca's arranger Mantovani to re-score the music a fourth lower. Piano transposition was more frequent, though in my experience only for vocal records. Even so, it happened less often than may be supposed. Accompanist Gerald Moore related how he would rehearse a difficult song transposed down a semitone, but play in the right key on the actual take, forcing his singer to do it correctly. So transposition isn't common, and it's usually possible to detect when the voice-quality is compared with other records of the same artist. For the modern engineer the problem is to get sufficient examples to be able to sort the wheat from the chaff. A more insidious trap is the specialist producer who's been listening to the recordings of a long-dead artist all his life, and who's got it wrong from Day 1 !

A useful document is L. Heavingham Root's article "Speeds and Keys" published in the *Record Collector* Volume 14 (1961; pp. 30-47 and 78-93). This gives recommended playing-speeds for vocal Gramophone Company records during the "golden years" of 1901 to 1908, but unfortunately a promised second article covering other makes never appeared. Mr. Root gave full musical reasons for his choices. Although a scientific mind would challenge them because he said he tuned his piano to C = 440Hz (when presumably he meant A = 440Hz), this author has found his recommendations reliable.

Other articles in discographies may give estimated playing-speeds accurate to four significant figures. This is caused by the use of stroboscopes for measuring the musically-correct speed, which can be converted to revolutions-per-minute to a high degree of accuracy; but musical judgement can never be more accurate than two significant figures (about one percent), for reasons which will become apparent in the next few paragraphs. Tighter accuracy is only necessary for matching the pitches of two different recordings which will be edited together or played in quick succession.

5.6 Strengths and weaknesses of "standard pitch"

The next difficulty lies in ascertaining the pitch at which musical instruments were tuned. There has always been a tendency for musicians to tune instruments sharp for "extra brilliance", and there is some evidence that standard pitches have risen slowly but consistently over the centuries in consequence. There were many attempts to standardise pitch so different instruments could play together; but the definitive international agreement did not come until 1939, after over four decades of recording using other "standards." You will find it fascinating to read a survey done just prior to the International Agreement. Live broadcasts of classical music from four European countries were monitored and measured with great accuracy. (Ref. 4). There were relatively small differences between the countries, the averages varying only from A = 438.5 (England) to A = 441.2 (Germany). Of the individual concerts, the three worst examples were all solo organ recitals in winter, when the temperature made them go flat. When we discount those, they were overshadowed by pitch variations which were essentially part of the language of music. They were almost exactly one percent peak-to-peak. Then, as now, musicians hardly ever play exactly on key; objective measurement is meaningless on expressive instruments. Thus, a musically trained person may be needed to estimate what the nominal pitch actually is.

Other instruments, such as piano-accordions and vibraphones, have tuning which cannot be altered. When a band includes such instruments, everyone else has to tune to them, so pitch variations tend to be reduced. So ensembles featuring these instruments may be more accurate.

Instead of judging by ear, some operators may prefer the tuning devices which allow pop musicians to tune their instruments silently on-stage. Korg make a very wide range, some of which can also deal with obsolete musical pitches. They can indicate the actual pitch of a guitar very accurately, so it could be possible to use one to measure a recording. But they give anomalous results when there is strong vibrato or harmonics, so this facility must be used with care, and only on recordings of instruments with "fixed tuning" (by which I mean instruments such as guitars with frets, which restrict "bending" the pitch as a means of musical expression). In other cases (particularly ensembles), I consider a musical ear is more trustworthy.

5.7 Non-"standard" pitches

Before the 1939 Agreement, British concert pitch (called "New Philharmonic Pitch" or "Flat Pitch") was A = 435 at 60 degrees Fahrenheit (in practice, about 439 at concert-hall temperatures). The International Agreement rounded this up to A = 440 at 68 degrees (20 Celsius), and that was supposed to be the end of the matter. Nevertheless, it is known that the Berlin Philharmonic and the Philadelphia orchestras use A = 444Hz today (a Decca record-producer recalled that the Vienna Philharmonic was at this pitch in 1957, with the pitch rising even further in the heat of the moment). The pianos used for concertos with such orchestras are tuned even higher. This may partly be because the pitch goes up in many wind instruments as the temperature rises, while piano strings tend to go flatter. Concert-halls in Europe and America are usually warmer than 68 Fahrenheit; it seems that only us poor Brits try using 440Hz nowadays!

The next complication is that acoustic recording-studios were deliberately kept very warm, so the wax would be soft and easy to cut. Temperatures of 90F were not uncommon, and this would make some A=440 instruments go up to at least 450. On the other hand, different instruments are affected by different amounts, those warmed by human breath much less than others. (This is why an orchestra tunes to the oboe). The human voice is hardly affected at all; so when adjusting the speed of wind-accompanied vocal records, accurate voice-quality results from *not* adjusting to correct tuning pitch. Thus we must make a cultural judgement: what are we aiming for, correct orchestral pitch or correct vocal quality? For some types of music, the choice isn't easy.

There are sufficient exceptions to A = 440 to fill many pages. The most important is that British military bands and some other groups officially tuned to "Old Philharmonic Pitch" or "Sharp Pitch" before 1929, with A at a frequency which has been given variously as 452.5 and 454. Since many acoustic recordings have military band accompaniments, this shows that A = 440 can be quite irrelevant. Much the same situation occurred in the United States, but I haven't been able to find out if it applied elsewhere. Nor do I know the difference between a "military" band and a "non-military" one; while it seems British ones switched to A = 440 in about 1964. Before that, it was apparently normal for most British wind players to possess two instruments, a "low pitch" one and a "high pitch" one, and it was clearly understood which would be needed well in advance of a session or concert.

Of course, most ethnic music has always been performed at pitches which are essentially random to a European, and the recent revival of "original instrumental technique" means that much music will differ from standard pitch anyway.

Because of their location in draughty places, pipe organs tended to be much more stable than orchestras. Many were tuned to Old Philharmonic Pitch when there was any likelihood of a military band playing with them (the Albert Hall organ was a notorious example). Because an organ is quite impossible to tune "on the night", the speed of such location recordings can actually be set more precisely than studio ones; but for perfect results you must obviously know the location, the history of the organ, the date of the record, and the temperature!

With ethnic music, it is sometimes possible to get hold of an example of a fixed-pitch instrument and use it to set the speed of the reproduction. Many collectors of ethnic music at the beginning of the century took a pitch-pipe with them to calibrate the cylinder recordings they made. (We saw all sorts of difficulties with professionally-made cylinders at the start of section 5.4, and location-recordings varied even more widely). Unfortunately, this writer has found further difficulties here - the pitch of the pitch-pipe was never documented, the cylinder was often still getting up to speed when it was sounded, and I even worked on one collection where the pitch-pipe was evidently lost, some cylinders were made without it, and then another (different) one was found! In these circumstances, you can sometimes make cylinders play at the correct relative pitches (in my case, "nailed down" more closely by judging human voices), but you cannot go further. Sometimes, too, if you know what you're doing, you can make use of collections of musical instruments, such as those at the Horniman or Victoria & Albert museums.

5.8 The use of vocal quality

Ultimately, however, the test must be "does it sound right?" And with some material, such as solo speech, there may be no other method. Professional tape editors like myself

were very used to the noises which come from tape running at strange speeds, and we also became used to the effect of "everything clicking into place". One of the little games I used to play during a boring editing-session was to pull the scrap tape off the machine past the head, and try my luck at aiming for fifteen inches per second by ear, then going straight into "play" to see if I'd got it right. Much of the time, I had. All the theories go for nothing if the end-result fails to "gel" in this manner. Only when one is forced to do it (because of lack of any other evidence) should one use the technique; but, as I say, this is precisely when we must document *why* a solution has been adopted.

The operator must, however, be aware of two potential difficulties. The first is that the average human being has become significantly bigger during the past century. It is medically known that linear dimensions have increased by about five percent. (Weights, of course, have increased by the cube of this). Thus the pitch of formants will be affected, and there would be an even greater effect with the voices of (say) African pygmies.

John R. T. Davies also uses the "vocal quality" technique. He once tried to demonstrate it to me using an American female singer. He asserted that it was quite easy to judge, because at 78rpm the voice was "pinched", as if delivered through lips clenched against the teeth. I could hear the difference all right, but could not decide which was "right". It wasn't until I was driving home that I worked out why the demonstration didn't work for me. I am not acquainted with any native American speakers, and I am not a regular cinemagoer. So my knowledge of American speech has come from television, and we saw earlier that American films are transmitted four percent faster in Europe. So I had assumed that American vocal delivery was always like that. The point I'm making is that personal judgements can be useful and decisive; but it's vital for every individual to work out for himself precisely where the limits of his experience lie, and never to go beyond them.

Nevertheless I consider we should always take advantage of specialist experience when we can. When John R. T. Davies isn't certain what key the jazz band is playing in, he gets out his trumpet and plays along with the improvisation to see what key gives the easiest fingering. I sometimes ask a colleague who is an expert violinist to help me. She can recognise the sound of an "open string", and from this set the speed accurately by ear. And, although the best pianists can get their fingers round anything, I am told it is often possible to tell when a piano accompaniment has been transposed, because of the fingering. This type of "evidence", although indisputable, clearly depends upon specialist knowledge.

5.9 Variable-speed recordings

So far, we've assumed that a record's speed is constant. This is not always the case. On 78rpm disc records, the commonest problem occurs because the drag of the cutter at the outside edge of the wax was greater than at the inside edge, so the master-record tended to speed up as the groove-diameter decreased. I have found this particularly troublesome with pre-EMI Columbias, though it can crop up anywhere. It is, of course, annoying when you try to join up the sides of a multi-side work. But even if it's only one side, you should get into the habit of skipping the pickup to the middle and seeing if the music is at the same pitch. On the other hand, some types of performances (particularly unaccompanied singers and amateur string players) tend to go flatter as time goes by; so be careful. A technique for solving this difficulty was mentioned in paragraph 6 of section 1.6; that is,

to collect evidence of the performance of the disc-cutter from a number of sessions around the same date.

Until about 1940 most commercial recording lathes were weight-powered, regulated by a centrifugal governor like that on a clockwork gramophone. A well-designed machine would not have any excess power capability, because there was a limit to how much power could be delivered by a falling weight. The gearing was arranged to give just enough power to cut the grooves at the outside edge of the disc, while the governor absorbed little power itself. The friction-pad of a practical governor came on gently, because a sudden "on/off" action would cause speed oscillations; so, as it was called upon to absorb more power, the speed would rise slightly. By the time the cutter had moved in an inch or two, the governor would be absorbing several times as much power as it did at the start, and the proportion would remain in the governor's favour. So you will find that when this type of speed-variation occurs, things are usually back to normal quite soon after the start.

The correction of fluctuating speeds is a subject which has been largely untouched so far. Most speed variation is caused by defects in mechanical equipment, resulting in the well-known "wow" and "flutter." The former is slow speed variation, commonly less than twenty times per second, and the latter comprises faster variations.

Undoubtedly, the best time to work on these problems is at the time of transferring the sound off the original medium. Much "wow" is essentially due to the reproduction process, e.g. eccentricity or warpage of a disc. One's first move must be to cure the source of the problem by re-centering or flattening the disc.

Unfortunately it is practically impossible to cure eccentric or warped cylinders. The only light at the end of the tunnel is to drive the cylinder by a belt wrapped round *the cylinder* rather than the mandrel, arranged so it is always leaving the cylinder tangentially close to the stylus. (A piece of quarter-inch tape makes by far the best belt!) If the mandrel has very little momentum of its own, and the pickup is pivoted in the same plane, the linear speed of the groove under the pickup will be almost constant. But this will not cure wow if differential shrinkage has taken place.

Another problem concerns cylinders with an eccentric bore. With moulded cylinders the only "cure" is to pack pieces of paper between mandrel and cylinder to bring it on-centre. But for direct-cut wax cylinders, the original condition should be recreated, driving the mandrel rather than the surface (Ref. 5).

However, it is possible to use one source of wow to cancel another. For example, if a disc has audible once-per-revolution *recorded* wow, you may be able to create an equal-but-opposite wow by deliberately orienting the disc off-centre. This way, the phases of the original wow and the artificial wow are locked together. This relationship will be lost from any copy unless you invoke special synchronisation techniques.

It is often asked, "What are the prospects for correcting wow and flutter on a digitised copy?" I am afraid I must reply "Not very good." A great deal has been said about using computers for this purpose. Allow me to deal with the difficulties, not because I wish to be destructive, but because you have a right to know what will always be impossible.

The first difficulty is that we must make a conscious choice between the advantages and disadvantages. We saw in chapter 1 that the overall strategy should include getting the speed right *before* analogue-to-digital conversion, to avoid the generation of digital artefacts. Nevertheless it is possible to reduce the latter to any desired degree, either by having a high sampling-frequency or a high bit-resolution. So we can at least minimise the side-effects.

Correction of "wow" in the digital domain means we need some way of telling the processor what is happening. One way to do this objectively is to gain access to a constant frequency signal recorded at the same time, a concept we shall explore for steady-state purposes in the next section. But when the speed variations are essentially random the possibilities are limited, mainly because any constant frequency signal is comparatively weak when it occurs. To extract it requires sharp filtering, and we also need to ignore it when it is drowned by wanted sound. Unfortunately, information theory tells us we cannot detect rapid frequency changes with sharp filters. To make things worse, if the constant frequency is weak, it will be corrupted by background noise or by variations in sensitivity. Although slow wow may sometimes be correctable, I am quite sure we shall never be able to deal with flutter this way. I am sorry to be pessimistic, but this is a principle of nature; I cannot see how we shall ever be able correct random flutter from *any* constant frequency which happens to be recorded under the sound.

But if the wow or flutter has a *consistent* element, for example due to an eccentric capstan rotating twenty-five times per second in a tape recorder, then there is more hope. In principle we could tell the computer to re-compute the sampling-frequency twenty-five times per second and leave it to get on with it. The difficulty is "slippage." Once the recording is a fiftieth of a second out-of-sync, the wow or flutter will be doubled instead of cancelled. This would either require human intervention (implying subjectivism), or software which could distinguish the effect from natural pitch variations in the wanted sound. The latter is not inconceivable, but it has not yet been done.

The computer may be assisted if the digital copy bears a fixed relationship to the rotational speed of the original. Reverting to discs, we might record a once-per-revolution pulse on a second channel. A better method is some form of rigid lock - so that one revolution of a 77.92rpm disc always takes precisely 33957 samples, for example. (The US equivalent would be a 78.26rpm disc taking 33776 samples). This would make it easier for future programmers to detect cyclic speed variations in the presence of natural pitch-changes, by accumulating and averaging statistics over many cycles. So here is another area of development for the future.

Another is to match one medium to another. Some early LPs had wow because of lower flywheel effect at the slower speed. But 78rpm versions are often better for wow, while being worse for noise. Perhaps a matching process might combine the advantages of both.

In 1990 CEDAR demonstrated a new algorithm for reducing wow on digitised recordings, which took its information from the pitch of the music being played. Only slow wow could be corrected, otherwise the process would "correct" musical vibrato! Presumably this algorithm is impotent on speech, and personally I found I could hear the side-effects of digital re-sampling. But here is hope when it's impossible to correct the fault at source. Unfortunately, CEDAR did not market the algorithm.

I hope this discussion will help you decide what to do when the problem occurs. For the rest of this chapter, we revert to steady-state situations and situations where human beings can react fast enough.

5.10 Engineering evidence

Sometimes we can make use of technical faults to guide us about speed-setting. Alternating-current mains can sometimes get recorded - the familiar "background hum." In Europe the mains alternates at a nominal frequency of 50Hz, and in America the

frequency is 60Hz. If it is recorded, we can use it to compensate for the mechanical deficiencies of the machinery.

Before we can use the evidence intelligently, we must study the likelihood of errors in the mains frequency. Nowadays British electricity boards are supposed to give advance notice of errors likely to exceed 0.1 percent. Britain was fortunate enough to have a "National Grid" before the second World War, giving nationwide frequency stability (except in those areas not on 50Hz mains). Heavy demand would slow the generators, so they had to be speeded up under light demand if errors were not to build up in synchronous electric clocks. So the frequency might be "high" as well as "low." Occasional bombing raids during the second World War meant that isolated pockets of Britain would be running independently of the Grid, but the overall stability is illustrated by Reference 6, which shows that over one week in 1943 the *peak* error-rate was only 15 seconds in three hours, or less than 0.15 percent. (There might be worse errors for very short periods of time, but these would be distinctly uncommon). After the war, the Central Electricity Generating Board was statutorily obliged to keep its 50Hz supplies within 2Hz in 1951 and 0.5Hz in 1971. My impression is that these tolerances were extremely rarely approached, let alone exceeded. However there is ample evidence of incompetent engineers blaming "the mains" for speed errors on their equipment.

An anecdote to illustrate that things were never quite as bad as that. In the years 1967-1971 I worked on a weekly BBC Radio programme lasting half-an-hour, which was recorded using A.C. mains motors on a Saturday afternoon (normally a "light current load" time), and reproduced during Monday morning (normally a "heavy load time," because it was the traditional English wash-day). The programmes consistently overran when transmitted, but only by ten or fifteen seconds, an error of less than one percent even when the cumulative errors of recording *and* reproduction were taken into account. In over twenty-five years of broadcasting, I never came across another example like that. However, I have no experience of mains-supplies in other countries; so I must urge you to find the tolerances in other areas for yourself.

We do not find many cases of hum on professional recordings, but it is endemic on amateur ones, the very recordings most liable to speed errors. So the presence of hum is a useful tool to help us set the speed of an anomalous disc or tape; it can be used to get us "into the right ballpark", if nothing else. This writer has also found a lot of recorded hum on magnetic wire recordings. This is doubly useful; apart from setting the "ballpark" speed, its frequency can be used to distinguish between wires recorded with capstan-drive and wires recorded with constant-speed takeup reels. But here is another can-of-worms; the magnetic wire itself forms a "low-reluctance" path for picking up any mains hum and carrying it to the playback head. It can be extremely difficult to hear one kind of hum in the presence of the other.

Portable analogue quarter-inch tape-recorders were used for recording film sound on location from the early 1960s. Synchronisation relied upon a reference-tone being recorded alongside the audio, usually at 50Hz for 50Hz countries and 60Hz for 60Hz countries. Back at base, this pilot-tone could be compared with the local mains frequency used for powering the film recording machines, so speed variations in the portable unit were neutralised. In principle it might be possible to extract accidental hum from any recording and use it to control a playback tape-recorder in the same way. This is another argument in favour of making an "archive copy" with warts and all; the hum could be useful in the future.

We can sometimes make use of a similar fault for setting the speed of a television soundtrack recording. The "line-scan" frequency of the picture sometimes gets recorded

amongst the audio. This was 10125Hz for British BBC-1 and ITV 405-line pictures until 1987; 15625Hz for 625-line pictures throughout the world (commencing in 1963 in Britain); 15750Hz for monochrome 525-line 30Hz pictures, and 15734.25Hz for colour NTSC pictures. These varied only a slight amount. For example, before frame-stores became common in 1985, a studio centre might slew its master picture-generator to synchronise with an outside broadcast unit. This could take up to 45 seconds in the worst possible case (to achieve almost a full frame of slewing). Even so, this amounts to less than one part in a thousand; so speed-setting from the linescan frequency can be very accurate. In my experience such high frequencies are recorded rather inefficiently, and only the human ear can extract them reliably enough to be useful; so setting the speed has to be done by ear at present.

Although my next remark refers to the digitisation procedures in Chapter 2, it is worth noting that the embedding of audio in a digital video bitstream means that there must be exactly 1920 digital samples per frame in 625-line television, and exactly 8008 samples per five frames in NTSC/525-line systems.

The 19kHz of the stereo pilot-tone of an FM radio broadcast (1st June 1961 onwards in the USA, 1962 onwards in Britain) can also get recorded. This does not vary, and can be assumed to be perfectly accurate - provided you can reproduce it.

John Allen has even suggested that the ultrasonic bias of magnetic tape recording (see section 9.3) is sometimes retained on tape well enough to be useful. (Ref. 7). We usually have no idea what its absolute frequency might be; but it has been suggested that variations caused by tape speed-errors might be extracted and used to cancel wow and flutter. I have already expressed my reasons why I doubt this, but it has correctly been pointed out that, provided it's above the level of the hiss (it usually isn't), this information should *not* be thrown away, e. g. by the anti-aliasing circuit of a digital encoder. Although it may be necessary to change the frequency down and store it on a parallel track of a multi-channel digital machine, we should do so. Again, it's a topic for the future; but it seems just possible that a few short-term tape speed variations might be compensated objectively one day.

There is one caveat I must conclude with. The recording of ultrasonic signals is beset with problems, because the various signals may interfere with each other and result in different frequencies from what you might expect. For example, the fifth harmonic of a television linescan at 78125Hz might beat with a bias frequency of 58935Hz, resulting in a spurious signal at 19190Hz. If you did not know it was a television soundtrack, this might be mistaken for a 19kHz radio pilot-tone, and you'd end up with a one percent speed error when you thought you'd got it exactly right. So please note the difficulties, which can only be circumvented with experience and a clear understanding of the mechanics.

5.11 Timings

There's a final way of confirming an overall speed, by timing the recording. This is useful when the accompanying documentation includes the supposed duration. Actually, the process is unreliable for short recordings, because if the producer was working with a stopwatch, you would have to allow for reaction-time, the varying perception of decaying reverberation, and any "rounding errors" which might be practised. So short-term timings would not be reliable enough. But for longer recordings, exceeding three or four minutes,

the documentation can be a very useful guide to setting a playing-speed. The only trouble is that it may take a lot of trial-and-error to achieve the correct timing.

I hope this chapter will help you to assess the likelihood, quantity, and sign of a speed error on a particular recording. But I conclude with my plea once again. It seems to me that the basis of the estimate should also be documented. The very act of logging the details forces one to think the matter through and helps against omitting a vital step. And it's only right and proper that others should be able to challenge the estimate, and to do so without going through all the work a second time.

REFERENCES

1: anon., "Lexiconning" (article), *Sight and Sound* Vol. 57 No. 1 (Winter 1987/8), pp. 47-8. It should be noted this article's main complaint was a film speeded by *ten* percent. A Lexicon would be essential to stop a sound like "Chipmunks" at this rate, although the same technology could of course be used for a four-percent change.

2: Friedrich Engel (BASF), Letter to the Editor, *Studio Sound*, Vol. 28 No. 7 (July 1986), p. 147.

3: John C. Fesler, London: *Hillandale News*, No. 125 (April 1982), p. 21.

4: Balth van der Pol and C. C. J. Addink, "Orchestral Pitch: A Cathode-Ray Method of Measurement during a Concert" (*article*), *Wireless World*, 11th May 1939, pp. 441-2.

5: Hans Meulengracht-Madsen, "On the Transcription of Old Phonograph Wax Records" (*paper*), J.A.E.S., Jan/Feb 1976.

6: H. Morgan, "Time Signals" (Letter to the Editor), *Wireless World*, Vol. L No. 1 (January 1944), p. 26.

7: John S. Allen, "Some new possibilities in audio restoration," (*article*), ARSC Journal, Volume 21 No. 1 (Spring 1990), page 44.

6 Frequency responses of grooved media

6.1 The problem stated

The subject of this chapter raises emotions varying from deep pedantry to complete lack of understanding. Unfortunately, there has never been a clear explanation of all the issues involved, and the few scraps of published material are often misquoted or just plain wrong, whilst almost-perfect discographical knowledge is required to solve problems from one recording-company to the next. Yet we need a clear understanding of these issues to make acceptable "service copies", and we are compelled to apply the lessons rigorously for "objective copies". (My research shows we now have the necessary level of understanding for perhaps seventy-five percent of all grooved media before International Standards were developed). But for the "warts-and-all" copy, it isn't an issue of course.

Grooved disc records have never been recorded with a flat frequency response. The bass notes have always been attenuated in comparison with the treble, and when electrical methods are used to play the records back, it is always implied that the bass is lifted by a corresponding amount to restore the balance. You may like to demonstrate the effect using your own amplifier. Try plugging a good quality microphone into its "phono" input, instead of a pickup-cartridge. If you do, you will notice a boomy and muffled sound quality, because the phono circuitry is performing the "equalisation" function, which will not happen when you use a "Mic" input.

The trouble is that the idea of *deliberately* distorting the frequency response only took root gradually. In the days of acoustic recording (before there was any electronic amplification), it was a triumph to get anything audible at all; we shall be dealing with this problem in Chapter 11. Then came the first "electrical recording" systems. (I shall define this phrase as meaning those using an electronic amplifier somewhere - see Ref. 1 for a discussion of other meanings of the phrase, plus the earliest examples actually to be published). At first, these early systems were not so much "designed", as subject to the law of "the survival of the fittest." It was some years before objective measurements helped the development of new systems.

This chapter concentrates on electrical recordings made during the years 1925 to 1955, after which International Standards were supposed to be used. I shall be showing equalisation curves the way the discs were *recorded*. If you are interested in restoring the sound correctly, you will have to apply equalisation curves which are the *inverse* of these; that is, the bass needs to be *boosted* rather than *cut*.

The major reason for the importance of this issue is different from the ones of restoring the full "power-bandwidth product" that I mentioned in Chapter 1. Incorrect disc equalisation affects sounds right in the middle of the frequency range, where even the smallest and lowest-quality loudspeaker will display them loud and clear – usually at a junction between a modern recording and an old one. The resulting "wooliness" or "harshness" will almost always seem detrimental to the archived sound.

6.2 A broad history of equalisation

Electrical recording owed its initial success to the Western Electric recording system. Although this was designed using scientific principles to give a "flat frequency response," it had at least one undefined bass-cut which needs correction today, and other features if we are ever to achieve "high fidelity" from its recordings. So its success was partly accidental. The recording equipment dictated the equalisation, rather than the other way round.

During the next twenty years the whole process of making an acceptable record was a series of empirical compromises with comparatively little scientific measurement. During the Second World War accurate methods of measurement were developed, and after the war the knowledge of how to apply these to sound reproduction became more widely known. Thus it became possible to set up "standards", and modify equipment until it met those standards. Thus professionals (and, later, hi-fi fanatics) could exchange recordings and know they would be reproduced correctly.

This last phase is particularly troublesome. There were nine "standards" which users of disc recording equipment were invited to support between 1941 and 1953, and the ghastly details will be listed in sections 6.62 onwards. If you put your political thinking-cap on, and conclude that such chaos is typical of a Free Market Economy, I reply that State Monopolies could be just as bad. For example, between 1949 and 1961 the British Broadcasting Corporation had three "standards" used at once, none of which were International ones!

Most record manufacturers had different recipes which we can describe in scientific language. The number of recipes isn't just because of the "Not Invented Here" syndrome, but there was at least one manufacturer who kept his methods a trade secret because he feared his competitive advantage would be harmed! Two international standards were established in 1955, one for coarsegroove records and one for microgroove records. The latter has several names, but most people call it by the name of the organisation which promoted it, the Recording Industries Association of America. It is on any present-day "Phono Input," and I shall call it "RIAA" from now on.

So if you are interested in the faithful reproduction of pre-1955 records, you should at least know that an "equalisation problem" may exist.

6.3 Why previous writers have gone wrong

This section is for readers who may know something already. It summarises three areas in which I believe previous writers have got things wrong, so you can decide whether to read any more.

(1) Equalisation is largely independent of the make of the disc. It depends only upon *who cut the master-disc and when*. (I shall be using the word "logo" to mean the trademark printed on the label, which is something different again!) I'm afraid this implies you should be able to detect "who cut the master-disc and when" by looking at the disc, not the logo. In other words, you need discographical knowledge. I'm afraid it's practically impossible to teach this, which may explain why so many previous writers have made such a mess of things.

(2) It is best to define an equalisation curve in unambiguous scientific language. American writers in particular have used misleading language, admittedly without committing *gross*

errors along the way. I shall be using "microseconds", and shall explain that at the end of section 6.7 below.
(3) The names of various "standards" are themselves ambiguous. For instance, when International Standards became operational in 1955, most old ones were re-named "the new Bloggs characteristic" or words to that effect. I recently found a microgroove nitrate dated 24-1-57 whose label bore the typed message: "Playback: New C.C.I.R., A.E.S., Orthoacoustic." (This was clearly RIAA, of course!) Similar considerations apply to curves designed for one particular format (for example, American Columbia's pioneering long-playing disc curve of 1948), which may be found on vintage pre-amplifiers simply called "LP" only - or worse still "Columbia" only - when neither name is appropriate, of course.

6.4 Two ways to define "a flat frequency response"

Equalisation techniques are usually a combination of two different systems, known for short as "constant velocity" and "constant amplitude." The former, as its name implies, occurs when the cutting stylus vibrates to and fro at a constant velocity whatever the frequency, provided the volume remains the same. This technique suited an "ideal" mechanical reproducing machine (not using electronics), such as the Orthophonic Victrola and its HMV equivalent gramophone of 1926. These scientifically-designed machines approached the ideal very closely. On such records, as the frequency rises the amplitude of the waves in the grooves falls, so the high frequencies are vulnerable to surface noise. On the other hand low frequencies cause high amplitudes, which have the potential for throwing the needle out of the groove (Fig. 1a). Thus all disc records are given some degree of bass cut compared with the idealised constant-velocity technique.

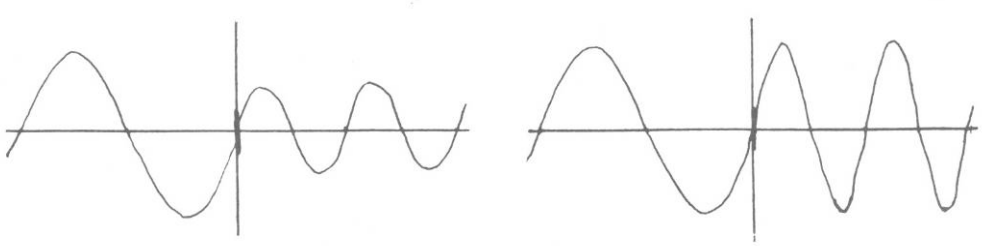


Fig 1a. Constant Velocity waveshape Fig 1b Constant Amplitude waveshape
These two diagrams depict how two musical notes, the second an octave above the first, would be cut onto lateral-cut discs using these two different systems.

Constant-amplitude recording overcomes both these difficulties. Given constant input, if varying frequencies are cut, the *amplitude* of the waves in the groove stays the same (Fig. 1b). Thus the fine particulate matter of the record is always overshadowed and the hiss largely drowned, while the low notes are never greater than the high notes and there is less risk of intercutting grooves. Unfortunately, the result sounded very shrill upon a clockwork gramophone.

Most record-companies therefore combined the two systems. In the years 1925-1945, most European record companies provided constant-velocity over most of the

frequency range to give acceptable results on acoustic gramophones, but changed to constant-amplitude for the lower frequencies (which were generally below the lower limit of such machines anyway) to prevent the inter-cutting difficulty. The scene was different in America, where the higher standard of living encouraged electrical pickups and amplifiers, and it was possible to use a greater proportion of constant-amplitude thanks to electronic compensation. From the mid-1930s, not only did many record manufacturers use a higher proportion of constant-amplitude, but another constant-amplitude section may have been added at high frequencies, which is usually called "pre-emphasis." More high-frequency energy is recorded than with the "constant-velocity" system. Thus the wanted music dominates hiss and clicks, which are played back greatly muffled without the music being touched. An equivalent process occurs today on FM broadcasts, TV sound, and some digital media.

In principle, magnetic pickups (not crystal or ceramic ones) give a constant voltage output when playing constant-velocity records. But this can be converted to the equivalent of playing a constant-amplitude record by applying a treble cut (and/or a bass boost) amounting to 6 decibels per octave. This can be achieved in mono with just two simple components - a resistor and a capacitor - so it is a trivial matter to convert an electronic signal from one domain to the other.

What exists on most electrically-recorded discs can be defined by one or more frequencies at which the techniques change from constant-amplitude to constant-velocity, or vice versa. "Phono" equalisers are designed to apply 6dB/octave slopes to the appropriate parts of the frequency spectrum, so as to get an overall flat frequency response. And graphs are always drawn from the "velocity" viewpoint, so constant-velocity sections form horizontal lines and constant-amplitude sections have gradients.

If you wish to reproduce old records accurately, I'm afraid I shan't be giving you any circuit diagrams, because it depends very much on how you propose to do the equalisation. Quite different methods will be needed for valves, transistors, integrated circuits, or processing in the digital domain; and the chronology of the subject means you will only find circuits for valve technology anyway. Personally, I don't do any of those things! I equalise discs "passively," using no amplification at the equalisation stage at all; but this implies neighbouring circuitry must have specific electrical impedances. This technique automatically corrects the relative phases (section 2.11), whether the phase changes were caused by acoustic, mechanical, or electronic processes in the analogue domain.

Next we have the problem of testing such circuitry. Over the years, many record companies have issued Frequency Test Discs, documenting how they intended their records to be reproduced. (Sometimes they published frequency response graphs with the same aim, although we don't actually know if they were capable of meeting their own specifications!). Such published information is known as "a Recording Characteristic," and I follow Terry's definition of what this means (Ref. 2): "The relation between the R.M.S electrical input to the recording chain and the R.M.S velocity of the groove cut in the disc." This gives rise to the following thoughts.

6.5 Equalisation ethics and philosophy

In the late 'twenties, a flat frequency response seems to have been the dominant consideration in assessing fidelity. Regardless of any other vices, a piece of equipment with a wide flat frequency range was apparently described as "distortionless." (Ref. 3).

Unless there is definite evidence to the contrary, we should therefore assume that 1920s engineers wanted their recordings to be reproduced to a flat frequency response, with the deficiencies of their equipment reversed as far as possible. This would certainly be true until the mid-thirties, and is sometimes true now; but there is a counter-argument for later recordings.

I mentioned how empirical methods ruled the roost. At the forefront of this process was the session recording engineer, who would direct the positions of the performers, choose a particular microphone, or set the knobs on his control-desk, using his judgement to get the optimum sound. As archivists, we must first decide whether we have the right to reverse the effects of his early microphones or of his controls.

The engineer may have had two different reasons behind his judgements, which need to be understood. One is an aesthetic one - for example, using a cheap omnidirectional moving-coil microphone instead of a more expensive one on a piano, because it minimises thumping noises coming from the sounding-board and clarifies the "right hand" when it is mixed with other microphones. Here I would not advocate using equalisation to neutralise the effects of the microphone, because it was an "artistic" judgement to gain the best overall effect.

The other is to fit the sound better onto the destination medium. In 78rpm days, the clarity would perhaps be further enhanced to get it above the hiss and scratch of the shellac. For reproduction of the actual 78 this isn't an issue; but if you are aiming to move the sound onto another medium, e.g. cassette tape or compact disc, it might be acceptable to equalise the sound to make it more faithful, so it suits today's medium better. (After all, this was precisely how the original engineer was thinking). This implies we know what the original engineer actually did to fit the sound onto his 78. We either need industrial archaeology, or the original engineer may be invited to comment (if he's available). I am very much *against* the principle of one lot of empirical adjustments being superimposed on someone else's empirical adjustments.

Personally I take the above argument to its logical conclusion, and believe we should only compensate for the microphone and the controls if there were no satisfactory alternatives for the recording engineer. Thus, I would compensate for the known properties of the 1925 Western Electric microphone, because Western Electric licencees had only one alternative. It was so conspicuously awful that contemporary engineers almost never used it. But listeners soon discovered that the silky sound of Kreisler's violin had been captured more faithfully by the acoustic process, despite the better Western Electric microphone being used. Subsequent measurements discovered the reason for its "acid sound." Therefore I consider it reasonable to include the effects of this microphone in our equalisation. It is debatable whether the microphone counts as part of the "recording chain" in Terry's definition of a "recording characteristic" - it wouldn't be considered so today, because of the engineer's conscious judgements - but when there were no alternatives, I think it may be reasonable to include it.

6.6 Old frequency records as evidence of characteristics

The concept of Frequency Records providing hard evidence of how old recording equipment performed is a modern idea, not considered by engineers in the past. Their concern was (usually) for calibrating reproducing equipment, not to provide us with concrete evidence of their weaknesses! Making test records is something this writer has attempted, and I can tell you from experience it isn't easy. It is one thing to claim a flat

frequency response to 15kHz; it is quite another to cut a flat frequency record to prove it. History shows that several "kludges" were adopted to achieve such records with the technology of the time. I shall be pointing out some of these kludges, and indicating where they might mislead a modern playback operator.

Given a frequency record, does it document the objective performance of the equipment, or does it document what the manufacturer fondly hoped his equipment should be doing? To use the word I employed above, do we know whether the disc has been "kludged" or not? I have stumbled across several ways of answering this question.

The first is to study the evolving performance of professional recording equipment through history, and put the frequency record in its chronological context. (*Could* a machine have cut 10kHz in 1944?). The next step is to study other sound recordings made by such equipment. (Do any of *them* carry frequencies as high as 10kHz?).

Another guide is to look at the physical features of the frequency disc. Does it appear that it was intended to be *used* as a standard of reference? If it carries a printed label, there would presumably be many copies, and that implies it was meant to be used; if someone has taken the trouble to label the frequencies by scrolls or announcements, that implies the disc was meant to be used; and if we have *two* contemporary records made by the same manufacturer and one is technically better than the other, then there is a higher probability that the "good" one was meant to be used. Thus we can say that the one intended to be *used* as a frequency disc is more likely to have been "kludged" to make it fit an intended characteristic and render it relatively "future-proof," while the others are more likely to document the actual everyday performance of the machine.

Many frequency records which were "meant to be used" carry tones at discrete frequencies. Unless the label says something specifically to the contrary, these cannot be used to assess the frequency response of the recorder, since there is no evidence that someone hasn't made adjustments between the different frequencies. If, however, the documentation makes it clear that the frequencies are recorded to some characteristic, it is possible to plot them on a graph (joining up several sides in the process if necessary) and achieve an overall frequency curve (or the *intended* frequency curve) for the recording-machine.

Even better evidence comes from a "sweep" frequency run, when a variable oscillator delivers frequencies covering a considerable range in a continuous recording. However, this too can be "cheated". A "sweep" recorded under laboratory conditions might not be representative of the hurly-burly of practical recording life, although it might very well document the *intended* performance. Other kludges were sometimes made to a machine to make it record to a particular characteristic, and we shall meet at least one definite example of such cheating in Section 6.11.

6.7 Two common characteristics

I shall now talk about the two most important equalisations I consider you need for making "service copies". The majority of European disc records made between 1925 and about 1952 had constant-velocity above about 300Hz and constant-amplitude below about 300Hz. This shape is known conversationally as a "Blumlein shape", after the EMI engineer who made several significant sound recording inventions. (He employed, but did not actually *invent*, this shape). The exact turnover frequency may vary somewhat, for example it might be 250Hz. Restoration operators call this "Blumlein 250Hz", even though the record in question was not cut with Blumlein's equipment, and there may be

no evidence that Blumlein ever used this particular frequency. It is a useful shorthand expression to describe the *shape*, nothing more. (Fig. 2)

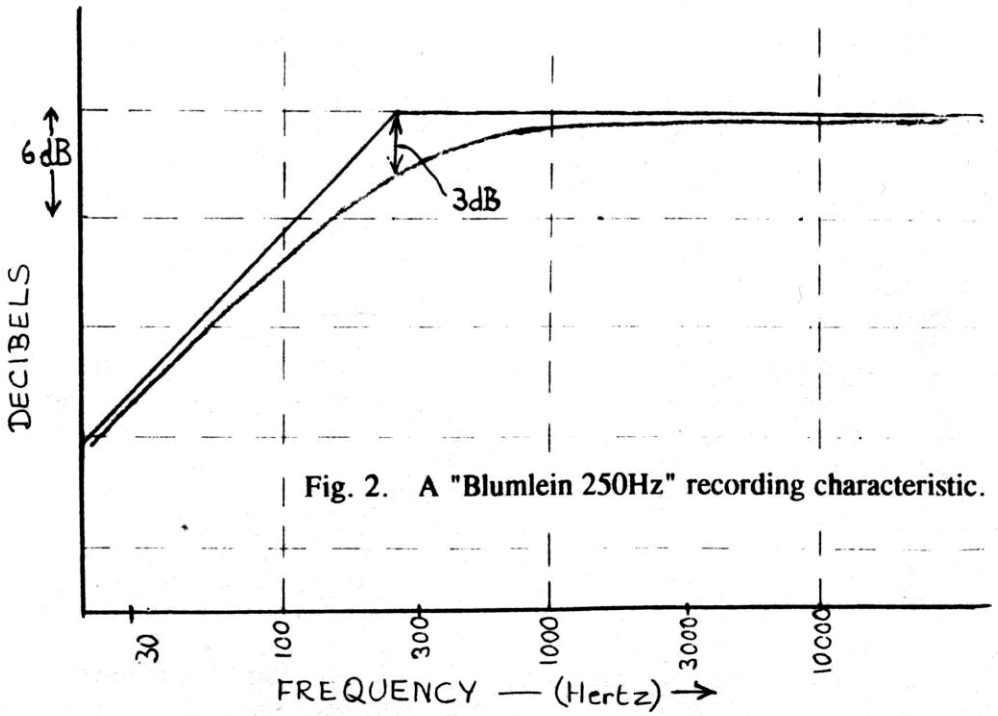


Fig. 2. A "Blumlein 250Hz" recording characteristic.

Figure 2.

The other important equalisation is the International Microgroove or "RIAA" one, which you probably have anyway. British citizens could find the 1955 version in British Standard 1928: 1955. This dealt with both coarsegroove and microgroove records. Since then there have been a number of revisions, and the current standard is BS 7063: 1989 (which is the same as IEC 98: 1987). This has one minor amendment to the original microgroove characteristic. I shall try to describe them both in words.

The international coarsegroove characteristic retained constant-velocity between 300Hz and 3180Hz, so most of the range of acoustic gramophones remained on the ideal curve, but the sections from 50 to 300Hz and from 3180Hz upwards were made constant-amplitude. But because it dates from 1955 when the coarsegroove format was nearly dead, and the subject matter was nearly always mastered on tape first, you should hardly ever need it.

For international standard microgroove, the constant-velocity section ran only from 500Hz to 2120Hz, so more of the frequency range approached constant-amplitude, and surface noise was further diminished. This was actually a compromise between the practices of several leading record companies, all slightly different. But it was found this type of bass cut caused *too little* amplitude at extremely low frequencies, and turntable rumble was apt to intrude. So the 1955 standard specified constant-velocity for all frequencies *below* 50Hz. A further development occurred in 1971, when some American record companies were trying the dbx Noise Reduction System on discs (section 9.7). The standard was changed so there should be decreased levels below 25Hz, because it was found that unevenness in this region had dramatic side-effects with such noise reduction systems. But noise reduction never "took off" on disc records in the way that it did on

cassettes, and relatively few discs carried significant amounts of sound energy below 25Hz anyway.

Engineers therefore define nearly all characteristics in terms of intersecting straight lines with "flat" or "6dB/octave" sections. But in practice electronic circuits cannot achieve sharp corners, and everyone accepts that the corners will be rounded. The straight line ideals are called "asymptotes". If there are only two intersecting asymptotes, a practical circuit will have an output three decibels away at the corner, and this intersection is therefore known conversationally as the "three dB point." Thus we may say "This Blumlein shape has a 3dB point at 250Hz" (and you will see why by referring back to Fig. 2.

Another way of defining the intersection is to specify the ratio of resistance to reactance in an electronic circuit - in other words, to specify the actual circuit rather than its result. For reasons it would take too long to explain here, this method gives an answer in microseconds. (It's related to the time it would take for the reactance to discharge through the resistance). Earlier, I said the international microgroove curve changed from constant-velocity to constant-amplitude at 2120Hz. This corresponds to a time constant of 75 microseconds. These are just two different ways of saying the same thing.

You could, of course, say the same thing in yet a third way, for example "Recorded level = +13.5dB at 10kHz." In other words, you could describe the characteristic by its effect at some extreme frequency, rather than at the intersection of the asymptotes. You may frequently encounter this style of description (favoured in America); but I shan't be using it, because it's fundamentally ambiguous. Decibels are a method of defining the relation between *two* things, and if you're an engineer you immediately ask "plus 13.5dB with respect to *what*?" A mid-frequency tone, such as 1kHz? Or the asymptote (which, in many cases, is approached but never reached)? Or an idealised "low frequency" (in which case, what about the 500Hz bass-cut - do we "count" it, or don't we)? And so on. To eliminate the ambiguities, and to prove I've done my homework properly, I shall be describing equalisation curves *in microseconds only*.

6.8 Practical limits to equalisation

As I hinted above, asymptotes of 6dB/octave are easy to achieve in electronics; but I want to draw your attention to some practical limitations. A specification which says a characteristic should be constant-amplitude for all frequencies above 75 microseconds is, of course, impossible to achieve. Quite apart from the obvious mechanical limitations of a cutting stylus doubling its velocity every time the frequency doubles, the electronics cannot achieve this either. If the recording amplifier has to double its output every time the frequency doubles, sooner or later it is going to "run out of amplification." No practical equipment exists in which this type of asymptote can be achieved.

In practice, equipment was designed to be approximately correct throughout the audio frequency-range, and to terminate at ultrasonic frequencies. It is quite normal for a couple of decibels to separate the theoretical and practical characteristics under these conditions. Unfortunately the differences depend upon the exact compromises made by the designer, and with the exception of an equalisation system built by myself, I do not know of any that have ever been documented! They are probably not very significant to the listener. But you should remember it is easy to destroy an infinite number of decibels of sound during playback, while it is never possible to boost an infinite number of decibels

when recording, so there is sometimes an inherent asymmetry between the two processes.

This seems to be the occasion to say that digital techniques are (in general) unsuitable for reversing such equalisation. Slopes of 6dB's per octave imply responses up to infinity (or down to zero), and digital techniques are inherently incapable of handling infinities or zeros. And if they do, side-effects are generated which defeat the "anti-aliasing filter" and add high-frequency noises (as described in section 3.2), while relative phases (section 2.11) may not be corrected. If "infinite impulse response" algorithms should ever become available, I would consider them; otherwise it's one of the very few cases where it's actually better to convert a digital signal back to analogue, re-equalise it, and then convert it back to digital again.

6.9 Practical test discs

In effect, the next two sections will comprise reviews of different frequency records which I have found useful in calibrating pickups and equalisers. From Section 6.20 onwards I shall be dealing with their other role for a restoration operator, documenting the performance of obsolete recording machinery. But first I must stress the importance of having these tools. Do not assume that, simply because you have modern equipment, it will outperform old recording systems. It should be your duty, as well as a matter of courtesy, to make sure your equipment does justice to what former engineers have recorded for you. Most operators get a shock when they play an old test disc with a modern pickup. Some find themselves accusing the test disc of being inaccurate, so widely can the performance deviate from the ideal. Rogue test discs *do* exist; but I shall mention all the faults I know below, and later when we look at the "industrial archaeology."

The first thing to say is that any test disc can give the wrong results if it is worn. All other things being equal, wear is worst at high frequencies and high volumes. Thus I prefer test discs with low volumes. Responsible reviewers in hi-fi magazines always used a new disc to test each pickup undergoing review; but this is obviously very expensive. In practice you cannot afford to thrash a new test disc every time you change a stylus, but you may find it difficult to separate the effects of surface noise when measuring a used disc with a meter. The trick is to use an oscilloscope and measure the sinewave with a graticule. Another trick is to use a relatively sluggish meter, rather than a peak-reading instrument. If the high frequencies give consistent results with two or three playings with your standard pickup, they will give valid results for hundreds of such playings, even though background noise may accumulate.

6.10 International standard microgroove test discs

This is usually the most important type of test disc. Not only will the majority of work be to this standard, but it is the starting point from which one develops coarsegroove and non-standard microgroove methods. There are many such discs made commercially. However, many of the test discs in this section have been "kludged" by taking advantage of the 75-microsecond pre-emphasis of the International Standard curve. To cut the highest frequencies onto the master disc, a constant frequency would be fed into the cutter (at about 8kHz), and then the cutting-turntable would be deliberately *slowed*. The recorded amplitude remained unchanged, of course; and the limitations I expressed in

Section 6.8 (plus others such as stereo crosstalk) were circumvented. The discs below are all 33rpm 12-inch LPs unless stated otherwise.

Mono (lateral):

B.B.C. (U.K) FOM.2
Decca (U.K) LXT.5346
Decca (U.K) 45-71123 (7-inch 45rpm)
EMI (U.K) TCS.104
HMV (U.K) ALP.1599
Urania (USA) UPS.1

Stereo:

B & K (Denmark) QR.2009 (45rpm)
B & K (Denmark) QR.2010 * (both published by Brüel & Kjaer)
C.B.S (USA): STR.100, STR.120, STR.130, BTR.150
Decca (U.K) SXL.2057
DIN (Germany) 1099 111 and 1099 114 (published by Beuth-Vertrieb GmbH, Berlin)
DGG (Germany) DIN 45543
Electronics World Test Record #1 (USA) (this is a 7-inch 33rpm disc, useful for measuring inner-groove attenuation; the frequency tests are mono)
EMI (U.K) TCS.101 (constant frequency bands)
EMI (U.K) TCS.102 (gliding tone)
High Fidelity (Germany) 12PAL 3720
London (USA) PS121
Shure (USA) TTR.102 *
VEB (East Germany) LB.209, LB.210
Victor Company of Japan JIS.21 (spot frequencies)
Victor Company of Japan JIS.22 (sweep frequencies)

(*Items marked with asterisk comprise only "sweeps" for use with automatic plotters; the frequency range is swept too fast for manual logging of measurements).

Consumer demonstration records with frequency tests - mono:

Acos (U.K) un-numbered (7-inch 45rpm; one side is a re-pressing from Decca 45-71123 above)
Urania (USA) 7084, also on Nixa (U.K) ULP.9084. (Five frequencies only)
Vox (USA) DL.130

Consumer demonstration records with frequency tests - stereo:

Audix (U.K) ADX.301
BBC Records (U.K) REC.355 (the frequency test is mono only)
C.B.S (USA) STR.101
Sound Canada Magazine (Canada) un-numbered: mono "interrupted sweep"

6.11 Coarsegroove (78rpm) test discs

In my opinion the most important coarsegroove test disc is Decca K.1803, recorded in 1947 (which was also available in North America as London T.4997). This has one side with constant-velocity sweep from 14kHz to 3kHz with an accurately modulated post-war style V-bottomed groove pressed in shellac. It was in the general catalogue for many years, and when copies turn up on the second-hand market they should be pursued. It gives the constant-velocity section of the so-called "Blumlein" shape used by EMI, the world's largest recording company, between 1931 and 1953; and a similar characteristic was used by many other companies, including Decca itself until 1944.

Unfortunately, nothing's perfect. The other side covers the range from 3kHz downwards, and suffers from two faults. One is that the level is about 1.5dB higher than side 1. The other is that Decca made a "kludge" at the turnover frequency. Officially, there should be a -3dB point at 300Hz (531 microseconds); but the actual discs show this as a "zero point", not a "-3dB point." In other words, someone has adjusted the signals fed into the cutter so they follow the asymptotes of the curve, rather than the curve itself (Fig. 2). So you should expect a +3dB hump at 300Hz if you have the correct equaliser.

The EMI equivalent of this disc is HMV DB4037, cut in 1936. Its main advantage is that it is cut with a typical U-bottomed groove of the period, so it will show the effectiveness (or otherwise) of truncated elliptical styli. Being made of shellac and recorded at rather a low level, wear-and-tear do not seem to affect the measurements, so the fact that the record may be second-hand is not usually important.

Unfortunately there are many criticisms of DB4037. First, examination of the groove-walls by the Buchmann-Meyer method and by direct measurement by a stereo pickup shows that they each carry the correct level of modulation. But there are phase differences between the two groove walls, which make the monophonic output wrong at high frequencies; this was because the cutting stylus which cut the master wax for matrix 2EA3181-1 was inserted askew. The effect appears on quite a few published discs mastered with a Blumlein cutter, and the symptoms can be ameliorated by twisting the pickup cartridge in its headshell by up to thirty degrees.

DB4037 is also useless for exploring the extremes of the frequency range. Its upper limit is only 8500Hz, the low frequencies are swept rather fast, and Sean Davies has recently shown (*pers. comm*.) that there is a loss below about 100Hz. DB4037 is one record from a set of five which includes fixed-frequency tones, and the lowest tones happen to be on the back of DB4037; so one would think one could use these. Alas, no. The accompanying leaflet is vague and misleading, and it took me some time to work out what was going on. To sum up briefly, the fixed tones are cut to a "Blumlein 150Hz" characteristic (1061 microseconds), the sweep tone is cut to "Blumlein 500Hz" (318 microseconds), while the leaflet implies (but doesn't actually say) that the turnover frequency is 250Hz (636 microseconds). I hate to think how much confusion has been caused by this set of discs.

EMI deleted DB4037 after the war and substituted a more exotic disc - EMI JG449, which was also issued in Australia on two discs as His Master's Voice ED1189 and ED1190. It was made to the "Blumlein 250Hz" (636 microsecond) curve in the summer of 1948. It never appeared in any catalogue and is much rarer. Furthermore, it consists only of fixed frequencies, with (maddeningly) even-numbered kiloHertz on one side and odd-numbered kiloHertz on the other. So exploring resonances is very difficult; but as it

extends all the way up to 20kHz, and can be found pressed either in shellac or vinyl, it has a niche of its own.

Finally, I should mention Folkways (USA) FX-6100, issued in 1954 before international standards were established. It has a 78rpm side which seems to be "Blumlein 500Hz" (318 microseconds), although the label claims "0dB" at all frequencies. This is likely to be more accessible to American readers.

These coarsegroove discs are very important. Not only do we have a large number of records made with a "Blumlein" shaped characteristic or something similar, but it was often used by default for reasons we shall see in section 6.13. There is only one turnover, so it is easy to reverse-engineer it if subsequent research uncovers more information. And there is minimal effect upon the waveforms of clicks and pops, so a copy may subsequently be declicked using electronic techniques.

I consider the remaining discs in my review as "luxuries" rather than "necessities."

Decca K.1802 (or London T.4996 in the USA) is similar to K.1803, but recorded to the 78rpm "ffrr" characteristic - the world's first full-range recording characteristic, which included publicly-defined high-frequency pre-emphasis. This comprised a constant-amplitude section with a +3dB point at 6360Hz (25 microseconds). There are sufficient Decca-group 78s around for it to be well worthwhile getting this right; from 1944 to 1955 Decca's chief engineer Arthur Haddy was responsible for getting the full frequency range accurately onto discs using this system, and it's clearly incumbent upon us to get it off again.

Ideally, you should also have a test disc for the International Standard Coarsegroove Characteristic, but such test discs are very rare, because the standard was introduced in 1955 when the coarsegroove format was nearly dead anyway. The only ones I have found are BBC DOM86, EMI JGS81, and an American one, Cook Laboratories No. 10. The latter is made of clear red vinyl, which could be another variable in the equation. As far as I know, only a few private coarsegroove discs will not have equivalent microgroove versions or master-tapes with better power-bandwidth product.

One sometimes finds differences because a test disc is made of vinyl (which is compliant) instead of shellac (which isn't). This effect will vary with the pickup and stylus being used, and I would encourage serious students to quantify the differences attributable to the particular pickup, and if necessary make adjustments when playing records made of different materials. In my experience, shellac always gives consistent results, because it is many orders of magnitude stiffer than any modern pickup stylus; but to show the size of the problem, I shall mention one case where I had a vinyl and a shellac version of the same coarsegroove frequency test record (it was EMI JG.449). Using a Shure N44C cantilever retipped with a 3.5 thou x 1.2 thou truncated elliptical diamond, the discs were in agreement at all frequencies below 6kHz. But above this, the responses increasingly differed; at 8kHz, the vinyl was -3dB, and at 13kHz it was -7.5dB. These figures were 1dB worse when the playing-weight was increased to the optimum for shellac (about 8 grams). By repeating the comparisons at 33rpm, it was possible to show that this was a purely playback phenomenon, and I was not damaging the disc by pushing the vinyl beyond its elastic limits; but one wonders what would have happened with nitrate!

These three sections by no means exhaust the list of useful frequency test discs, but many others suffer from defects. They help to identify the performance of old recording machinery with some degree of objectivity, but are not recommended for routine alignment of modern replay gear. I shall therefore defer talking about them until we get to the industrial archaeology of obsolete recording equipment.

6.12 Generalised study of electromagnetic cutters

We now consider the characteristics of cutterheads. A cutterhead converts electrical waveforms into modulations in a groove, and if it isn't perfect it will affect the wanted sound. "Simple" cutterheads were used for all electrically recorded lateral-cut discs between about 1925 and 1949. They were still being used by amateurs and semi-professionals until the mid-1960s; but after these approximate dates, more complicated cutters with "motional feedback" took over. The specific performance of a cutterhead can largely be neutralised by motional feedback.

Early cutterheads performed the subliminal function of determining recording characteristics. Microphones and amplifiers were intended to have a response which was uniform with frequency. This was rarely achieved to modern standards, of course, but it *was* an explicit aim. However most cutterheads had a non-uniform response which was found to be an advantage. Different makes of cutterhead would give different sections with constant-velocity and constant-amplitude features, so studying the cutterheads enables us to assess the objective performance of a recording machine before predetermined characteristics were adopted. In the early 1940s American broadcasters decided to use predetermined characteristics for their syndicated programmes, so cutterheads had to be modified or electronically compensated to bring them into line with the theoretical ideal. These theoretical characteristics will be left until section 6.62 onwards; but some organisations (and most amateurs) did not bring their cutting techniques into line with any particular standard until many years later.

There are very few cutterheads which do not conform to the simple "Blumlein shape" outline. The principal exceptions besides the motional-feedback types are piezo-electric cutters (confined to US amateur recording equipment of the mid-1940s), the Blumlein system (which involved "tuning" its resonance, described in sections 6.29 to 6.34below), and the BBC feedback cutterhead (in which the feedback was "non-motional" - the electromagnetic distortions were cancelled, but not the armature motion). So far as I know, all the others follow the same performance pattern, whether they work on moving-iron or moving-coil principles.

There is another reason for mentioning the general characteristics of cutterheads. When we do not know the apparatus actually used for a record, we can get a general outline of the frequency characteristic which resulted, although we may not know precise turnover frequencies. This at least enables us to avoid inappropriate action, such as variable-slope equalisation when we should be using variable-turnover. But we may not be able to guarantee high fidelity unless future research brings new information.

6.13 Characteristics of "simple" cutterheads

All cutterheads have to be capable of holding a cutting tool firmly enough to withstand the stresses of cutting a groove, while the tool is vibrated with as much fidelity to the electronic waveform coming out of the amplifier as possible. To achieve the first aim, the cutter must be held in a stiff mounting if it is not to be deflected by the stresses. In practice this means the cutter (and the armature to which it is attached) has a mechanical resonance in the upper part of the audio frequency range. The fundamental resonant frequency always lies between 3 and 10kHz for the "simple" cutterheads in this section.

It isn't usually possible to define the resonant frequency precisely, because it will vary slightly depending on the type of cutting stylus. The early steel cutters for nitrate discs, for example, were both longer and more massive than later sapphires in duralumin shanks; this effect may de-tune a resonance by several hundred Hertz. Fortunately it is rarely necessary for us to equalise a resonance with great precision. Various ingenious techniques were used to damp a cutter's fundamental resonance, and the effects upon nearby frequencies were not great.

To deflect the cutter, electric current from the amplifier had to pass through a coil of wire. In moving-iron cutters, the current magnetised an armature of "soft iron" which was attracted towards, or repelled from, magnetic pole-pieces. (In this context, "soft iron" means *magnetically* soft - that is, its magnetism could be reversed easily, and when the current was switched off it died away to zero). In moving-coil cutters, the current caused forces to develop in the wire itself in the presence of a steady magnetic field from the pole-pieces. The resulting motion was therefore more "linear", because there was nothing which could saturate; but the moving-iron system was more efficient, so less power was needed in the first place, all other things being equal.

The pole pieces were usually energised from permanent magnets; but electromagnets were sometimes used when maximum magnetic strength was needed. This was particularly true in the 1920s and early 1930s, before modern permanent magnet materials were developed.

The efficiency of the cutter depended on the *inductance* of the coil, not its resistance. To put this concept in words, it was the interaction between the magnetic field of the flowing current and the steady magnetic field which deflected the cutter. The electrical *resistance* of the coil, which also dissipated energy, only made the coil get hot (like an electric fire, but on a smaller scale). Inductive impedance always increases with frequency, while resistance is substantially constant with frequency.

If the coil were made from comparatively coarse wire, it would have lower resistance in relation to its inductance. But however the coil was designed, there would inevitably be a frequency at which the resistance became dominant - usually at the lower end of the frequency range. Sounds of lower pitch would be recorded less efficiently, because most of the power heated the wire instead of propelling the cutter. The slope of the frequency response of the cutterhead would change on either side of the frequency at which the resistance equalled the inductive impedance. This difference was asymptotic to 6dBs per octave.

You should be aware that there were also two second-order effects which affected the turnover frequency. First, the output impedance of the amplifier: the lower this was, the less power was wasted in the output stages. Modern amplifiers always have a low output impedance, because they are designed for driving loudspeakers which work best this way. But in the 1920s and 1930s low output impedances were less common, and this could affect the turnover frequency by as much as thirty percent compared with a modern amplifier. The consistency of archival recordings isn't affected, but you should be aware of the difficulty if you try measuring an old cutterhead connected to a modern amplifier. Some meddling recordists working with pirated cutterheads went so far as to wire a variable series resistance between the amplifier and the cutterhead to control the shape of the bass response (Ref. 4). The other effect was the strength of the field electromagnet before permanent magnets were used. A weak field would, in effect, cause more power to be wasted in the coil resistance. Most of the time engineers sought maximum efficiency; but written notes were sometimes made of the field voltage. Both methods formed practical ways of reducing volume and cutting bass at the same time.

Meanwhile, Newton's laws of motion determined the performance of the armature/stylus mechanism. It so happened that when the natural resonance was at a high frequency, the response was constant-velocity in the middle of the frequency range, just what was wanted for an acoustic gramophone. As the frequency went up, the elasticity of the armature "pulled the stylus ahead of the magnetism," and the velocity tended to increase; but meanwhile the magnetism was falling because of the coil inductance. These effects neutralised each other, and the stylus moved with the same velocity when a constant voltage was applied. Thus, simple cutterheads gave constant-velocity between the effects of the coil resistance at low frequencies and the effects of the resonance.

All this can be made clearer by an actual example. The graph below (Fig. 3) documents a test I carried out many decades ago, to check the response of a "Type A" moving-iron cutterhead. (These were mounted on BBC Type C transportable disc cutters, which were carried round the country in the back of saloon cars between about 1938 and 1960). At low frequencies the bass is cut by the coil resistance; in this case, the coil has the somewhat unusual nominal impedance of 68 ohms so the bass-cut occurs at the desired frequency, 300Hz, which I allowed for on reproduction. The fundamental resonance of the armature is at 4kHz. Between these two frequencies, the output is essentially constant-velocity.

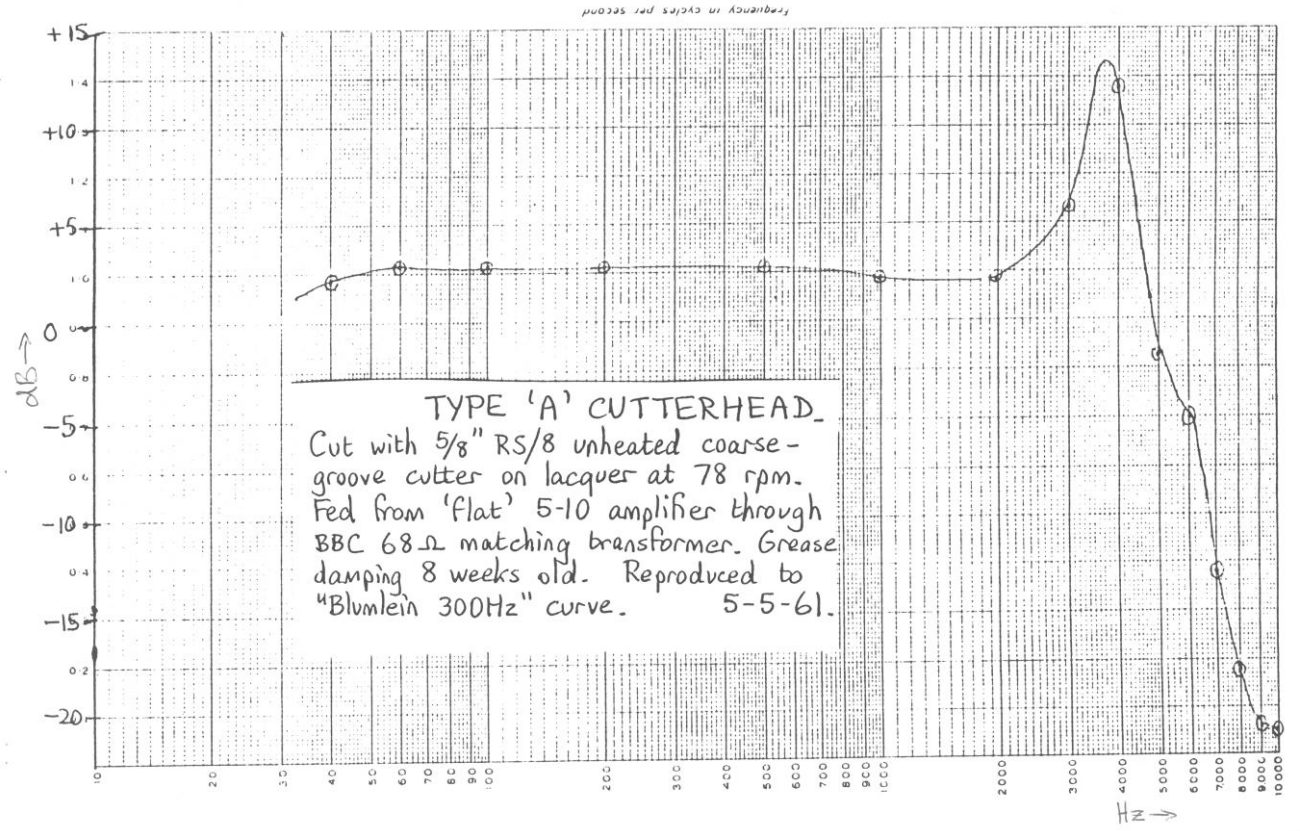


Figure 3

As you can see from the above graph, cutterheads may have a large peak in their output at their resonant frequency. It was this difficulty which was responsible for the relatively late arrival of electrical recording. Western Electric's breakthrough happened when they showed how the effect could be controlled mechanically. It was necessary to use a "resistive" element. I use this word to describe a mechanical part, as opposed to the

electrical property of wire I mentioned earlier. A "resistive element" has the property of absorbing energy equally well at all frequencies. With mechanical parts, a massive element tends to be easier to move at low frequencies, and a compliant element tends to be easier to move at high frequencies. At low frequencies where the mass is dominant, it tends to delay the motion, and there is a "phase lag." At high frequencies, where the compliance is dominant, the springiness tends to pull the armature ahead of the signal, and there is a "phase lead." Where the mass and compliance have equal effects, the phase lag and the phase lead cancel, and resonance occurs, resulting in more motion, apparently contradicting the laws of conservation of energy. Only mechanical resistance prevents infinitely fast motion.

The Western Electric people hung a long rubber tube on the armature, the idea being that the energy would be conducted away and dissipated in the rubber. The difficulty lay in supplying rubber with consistent properties. When the rubber line was new, it worked according to specification, and we do not need to take any special action to cancel a resonance today. But as the rubber aged, it gained compliance and lost resistance - it "perished." Top professional recording companies were able to afford new spare parts, but by about 1929 there was rather a lot of ill-understood tweaking going on in some places (Ref. 4). When this happened, the resonant frequency and the amplitude cannot be quantified, so current practice is simply to ignore the problem.

In the next couple of decades, many other types of resistive element were used. Fluid materials, such as oils and greases, had only resistive properties, not compliant ones. But grease could dry out, or refuse to remain packed against the armature where it should have been; Fig. 3 demonstrates the result. Oil could be used, soaked in paper inserted between armature and pole-pieces, kept there by surface-tension. Or the cutterhead might have a sump, with the cutting-tool poking out of the bottom through an oil-proof but flexible seal. Thixatropic greases such as "Viscaloid" became available after the war. Many amateur machines used something like conventional bicycle valve-rubber round the armature. Although this required renewal every few years, there was a sufficiently high ratio of resistance to compliance to make the resonance inaudible, although it could still be shown up by measurements.

Above the frequency of resonance, the stylus velocity would have fallen at a rate asymptotic to twelve decibels per octave, unless the armature was shaped so as to permit other resonances at higher frequencies. Such a resonance can just be discerned in Fig. 3, but it isn't audible. If the surface-noise permits (a very big "if"!), one could in principle equalise this fall, and extend the frequency range. I have tried experiments on these lines; but apart from being deafened by background noise like escaping steam, this often reveals lots of "harmonic distortion," showing that something was overloading somewhere (usually, I suspect, a moving-iron armature). Clearly, the original recording engineer was relying on the restricted frequency-range of his cutterhead to filter off overload-distortion. Unless and until someone invents a computer process to detect and cancel it, I think it's better to leave things the way they are.

Personally, I ignore fundamental resonances unless they are so clearly audible that they can be "tuned out" by ear. This is usually only possible when they occur at a relatively low frequency, within the normal musical range (say below 4kHz), which tended to happen with the cutterheads in the next section. Even so, I believe it is important to study many different records before quantifying and cancelling any resonance, so I can use the principle of "majority voting" to distinguish between the effects of the cutterhead, and specific features of musical instruments or voices.

6.14 High-resistance cutterheads

I cannot give industrial archaeology evidence for these cutterheads, because as far as I know there isn't much in the way of written history or surviving artefacts. As I do not have the quantitative information I shall be using in subsequent sections, I shall explain instead their qualitative principle of operation. They were confined to inexpensive discs; in Britain, the first electrical Edison-Bells, Sternos, Imperials, and Dominions.

The simple electromagnetic cutterhead I have just outlined had a transition from constant-amplitude to constant-velocity at the low frequency end of the scale. Such cutters had a relatively low electrical impedance (tens of ohms). Before transistors, this meant a matching transformer between the output valves and the coil. It was a bulky and expensive component, and until the mid-1960s it is no exaggeration to say that the output transformer was the weakest link in any power amplifier. By winding the cutterhead with a high-impedance coil (thousands of ohms), it could be coupled directly to the amplifier's output stage. But this brought its own complement of problems.

The most important was that the electrical resistance of the wire was dominant. Consequently, the cutter recorded constant-amplitude for much more of its frequency range. (Fig. 4)

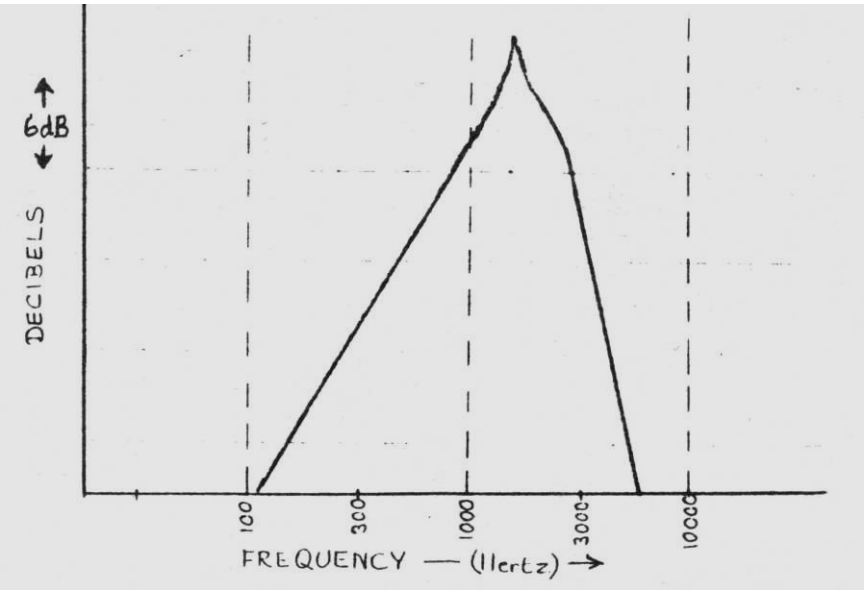


Figure 4. Response of typical high-resistance cutterhead.

Because more of the energy went into heating the coil rather than vibrating the cutter, such systems were less efficient. To compensate for this, the resonant frequency of the mechanism was lower so there was more output at mid-frequencies where the ear is most sensitive, and this resonance was less well damped.

In every case known to the author, the change from constant-amplitude to constant-velocity was *above* the frequency of resonance. Thus we have constant-amplitude through most of the musical range ending at about 1.5 or 2kHz with a massive resonance, then a falloff at 6dBs per octave until the resistive impedance equalled the inductive impedance, followed by a steeper fall at 12dBs per octave at the highest frequencies.

This sounds truly horrible today, and I mention it in case you need to deal with it. But I must also remind you that musicians would modify their orchestrations, their layouts, and their performing techniques to adapt to the strange circumstances (as they had done with acoustic recordings), because test records were judged on acoustic gramophones which gave better results with constant-velocity. All four makes cited above, for example, issued electrically-recorded dance-bands with brass bass lines. Thus it may not be fair to "restore the original sound", except perhaps with solo speakers, or solo musical instruments with which it was impracticable to alter the volumes of separate notes sounded together.

You may wonder how such a dreadful system came to be adopted. I suggest six reasons:

(1) It was no worse than acoustic recording, and when you add the extra advantage of electronic amplification, it was better.
(2) It was much cheaper than Western Electric's system (most things would have been!).
(3) It gave records which did not have large groove-swing amplitudes at low frequencies (section 6.4 above), and which therefore played on cheap gramophones without mistracking.
(4) Since most of the musical range was constant-amplitude, the advantages mentioned in section 6.4 were partly realised.
(5) The smaller companies did not have large sales among the affluent. The results matched the performance of down-market wireless sets.
(6) It enabled manufacturers to claim their records were electrically recorded, which was the great selling-point of the late 1920s!

It must also be said that at least two companies in Britain (British Homophone and Edison-Bell) seem to have altered their techniques after a year or so, to give something closer to a "Blumlein shape." Paul Voigt (of Edison Bell) later revealed that he had started with constant-amplitude (without explaining why, Ref. 5), and had to change to a compromise slope of 3dBs per octave, achieved by electronic compensation. He claimed he had been "the first to perceive the advantages of constant-amplitude"; but I believe this to have been an accidental discovery, since it was abandoned very quickly! Instead, he can take justified credit for being the first to use electronics to obtain a particular recording characteristic. In another case (Crystalate, makers of Imperial and Victory logos), Arthur Haddy was recruited as a recording engineer in 1929 because he said "I thought I could build a better lot on the kitchen table" (Ref. 6); but, so far as the London-recorded Imperials are concerned, the difference isn't apparent until late 1931 (a gradual changeover occurring between matrixes 5833 and 5913). The third case seems to be the Dominion Record Company, which seems never to have changed its practices before its bankruptcy in 1930.

6.15 Western Electric, and similar line-damped recording systems

For the next few sections, I shall consider the properties of the earliest successful disc recording systems using electronic amplifiers. They included cutterheads whose main resonance was damped by a "rubber line." This comprised a long rubber rod designed to

conduct vibrations away and dissipate them; but if the rubber went hard, vibrations might be reflected back into the armature instead of being totally absorbed.

The first successful cutterheads were actually of this type. They were based on research by the Bell Telephone Labs in the USA in the early 1920s, and were originally offered to the sound film market. Although some experimental "shorts" were made, film producers did not take to the idea immediately, although it was adopted by Vitaphone from 1926 onwards. The first commercial success was with the record industry, and the cutterhead formed part of a complete "system," comprising microphone, amplifier, cutterhead, and a scientifically-designed acoustical reproducer. Other record companies had similar cutterheads, and we will consider those as well.

6.16 Western Electric revolutionises sound recording

The technique was launched in 1925 by the manufacturing division of A. T & T, the Western Electric Company. It was almost immediately taken up by leading record makers, including the Victor/Gramophone Companies and the Columbia Companies on both sides of the Atlantic. The first published recordings date from mid-February 1925 (Ref. 1), and later developments continued in use until about 1949. All three components of the early equipment (the microphone, the amplifier, and the cutterhead) provide challenges to workers aiming to recover the original sound. First, we shall consider the earliest high-quality Western Electric microphone.

6.17 The Western Electric microphone

Fig. 5 shows the axial response of the Western Electric Type 361 condenser microphone (or "transmitter" as it was called in those days) (Refs. 7 and 8). The principal feature is a +7dB resonance at 2.9kHz due to the air in a cavity in front of the diaphragm. When the microphone was first invented, tests were done in an atmosphere of pure hydrogen; but the velocity of sound is much higher in hydrogen, and the cavity resonance was above the microphone's official limit. In air the defect is quite unambiguous. Such microphones were used as prime standards in acoustic laboratories for many years, and the defects had to be taken into account by the scientific community.

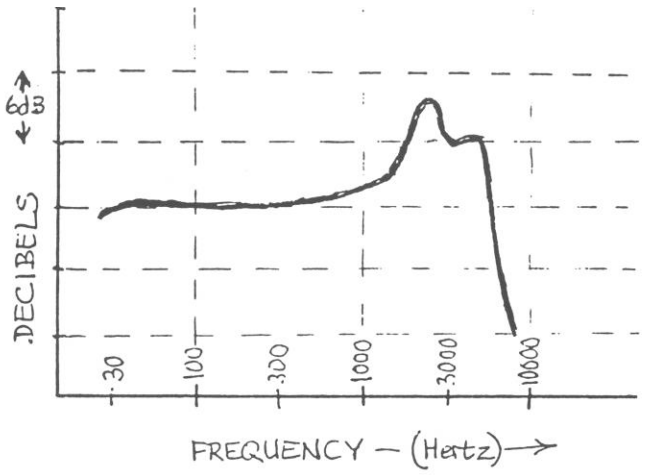


Fig. 5: Axial response of W. E. microphone.

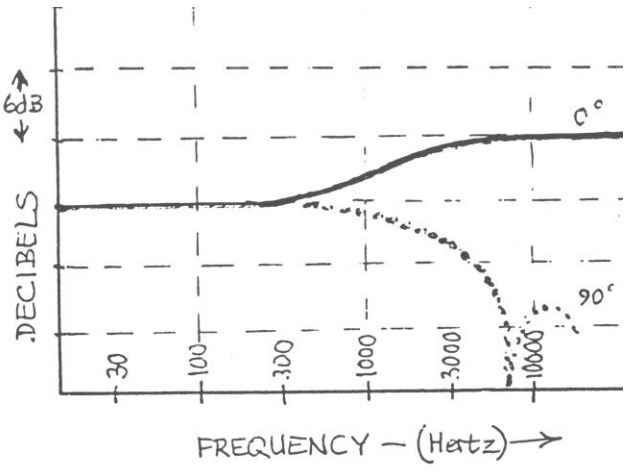


Fig. 6: Effects of directionality.

There was an increase at high frequencies for a second reason, which I have separated out to make Fig. 6. The microphone reflected some of the sound back at the performers, and as a consequence the diaphragm was subjected to increased sound pressures. When the performer was smack on axis, this action doubled the output at high frequencies - an increase of 6dBs - although other effects conspired to bring the extreme high-frequency response down again. But the dotted line in Fig. 6 shows that the microphone had less high-frequency output at the sides - ninety degrees off-axis. Since the performers might be anywhere in a semicircle around the microphone, the frequency response for any one individual might be anywhere between the solid and the dotted lines. It is normal practice to ignore the pressure-doubling effect altogether. Although we could compensate if we knew precisely where the artist was situated, we don't usually know this, and we certainly cannot treat performers differently if there were two or more at once. But personally, I do compensate for the main cavity resonance, because it reduces "mechanical quality" and surface noise at the same time.

The Model 394 seems to have been substituted for the Model 361 for gramophone work in 1929. Its capsule was the same, but the original preamplifier (in a wooden box) was replaced by a new design in a metal tube, like a modern condenser microphone. (Ref. 9). Although there may have been slight differences in the performance of the preamplifiers, and of the capsules (because they were re-housed), our current knowledge does not enable us to distinguish between them.

6.18 The Western Electric amplifier

The amplifier (which went between the microphone and cutterhead) had a substantially flat frequency response between 50Hz and 5kHz. Unfortunately, all examples seem to have been returned to the manufacturers in America, and although microphone capsules and cutterheads survive, there seems to be no electronic equipment. But volume controlling, in the manner known to recording operators today, was clearly possible from the outset. Many details are preserved in contemporary documentation, and may be useful to restoration operators. It seems the settings were logged because HMV submitted test pressings to a "wear test." Ideally, they had to survive 100 plays without wearing out; but if one failed, it had to be re-recorded using the same equipment at a lower volume.

Fred Gaisberg, the Gramophone Company's chief recording expert, has left an account which mentions another way of avoiding trouble (he was speaking of sudden *forte* timpani attacks): "The way to deal with these 'whacks' was to cut off the lower frequencies." (Ref. 10). And Moyer (Ref. 11) says "The actual effective low end of these curves is subject to some question, however, since it was common practice to use a rather elaborate "bass filter" to reduce the low-frequency response in order to obtain the best sound on average reproducers."

6.19 Documentation of HMV amplifier settings, 1925-1931

Settings of Western Electric equipment made by HMV may be found on the "Artists' Sheets", which may be consulted at the British Library Sound Archive. Microfilms 360 to 362 include British recordings and many International Red-Label artists. Recordings made

in Vienna and points east are on reels 385-6, Italy on 386-7, Spain on 387-8, Germany on 388, France on 388-9, and other places on 390-1. However, they do not show material recorded for the Zonophone logo.

It is my duty to point out that we have no independent description of the meanings of the settings. We must therefore "break the code." There is a stage between the breaking of a code and the acceptance of the decoded messages as historically accurate. The latter only comes when new material is decoded by new workers with consistent results. I am still at the intermediate stage.

The settings are in three or four columns at the right-hand side of the sheets. Because the columns are not labelled or numbered, I am forced to list them in the following manner.

EXTREME RIGHT-HAND COLUMN. Volume settings, going from L.1 to L.10 and H.1 to H.10. The letters are thought to refer to a switch on the microphone preamplifier offering "low" and "high" gain, while the numbers are thought to refer to ten taps on an auto-transformer after the cutting amplifier. (An eleventh was "Off.") The taps were equally-spaced, each step accounting for a uniform number of watts. These taps were originally provided for public-address applications, to control volume to several loudspeakers without wasting power. Often a *range* of taps is shown. (This information is deduced from a description of Western Electric public address amplifiers in Ref. 12). Occasionally there is a third parameter, generally a plain number covering a range from 4 to 16; this is thought to be a wire-wound volume control which could provide gradual changes in volume, in contrast to the switches.

SECOND COLUMN FROM THE RIGHT. Serial number of the microphone. When I say "serial number," I do not mean one allocated by Western Electric, but the Gramophone Company's number. Microphones in Continental studios often had a letter prefix (this is not always given); thus German recordings may be made with microphones numbered G.1 upwards; Milan, M.1 upwards.

THIRD COLUMN FROM THE RIGHT. "Serial number" of the amplifier or cutterhead (I do not know which).

FOURTH COLUMN FROM THE RIGHT. From May 1927 this column, which was originally designed for the copyright date of the music, was used for another purpose. It generally contains "2.L." or "3.L," but I have seen "1.L." on Italian artists' sheets. It is highly likely this documents what we now call a "brick-wall bass cut," and (judging from listening tests only) it seems "2.L." meant a cut below about 60Hz, and "3.L." below about 100Hz. These codes also appear on some Blumlein recordings (which we will consider in sections 6.29 to 6.34).

6.20 The Western Electric cutterhead

The 1925 version of Western Electric's cutterhead was upgraded in 1927, and the Gramophone Company of Great Britain distinguished the two by the terms "Western 1" and "Western 2", although these were not used by the manufacturer. Indeed, a Western Electric Recorder Type 2A using motional feedback was introduced in 1949. (By this time pre-determined recording characteristics were used. The cutterhead followed these fairly faithfully, so I shall leave the results until section 6.62 onwards). In order to circumvent the ambiguities of nomenclature, I have therefore decided to call the first manifestation "Western Electric 1A", the second "Western Electric 1B", and so on; although please note these names weren't used officially either!

The world's first frequency records were cut using "Western Electric 1A" equipment by the Victor Company in the United States. To judge by the matrix numbers which appear on the British pressings, they were done in two sessions in November 1925 and late January 1926. They were issued in the US on single-sided discs at rather high prices, and in Britain as a set of fifteen double-sided twelve-inch 78s in March 1929 (His Master's Voice DB1231 to DB1245). Unfortunately they comprise only fixed tones, there being no "sweep" which would enable us to assess frequency-response. They were used for many years by the magazine Wireless World for testing pickups. They are also quite rare (I have only ever found two), so I can't even use modern methods to measure what there is. But I am afraid my judgement is that we cannot get any useful information about the recorded frequency-response from these discs.

Fig. 7 reproduces two frequency curves, documenting the performance of the 1A and 1B electromagnetic cutters with their rubber line damping. They are redrawn from Figure 3 of the paper by Moyer (Ref. 11). The 1A version had a primary resonance at about 4.5kHz, above which its response would have fallen asymptotically to 12dB per octave. The model 1B extended the response slightly to about 5.5kHz. This seems to have been achieved by redesigning the means of pivoting the armature. The 1A had quite a lot of compliance, and an electronic click could "throw" the armature and cause it to stick to one of the pole pieces; the 1B tightened up the compliance so this happened less often and the fundamental resonant frequency was raised.

From section 6.13, you would have expected a 6dB/octave rolloff in the bass due to the resistance of the coil, and indeed there was; and when the rubber line was new, that is all there was. (Ref. 13). I have shown it by the dotted line in Fig. 7.

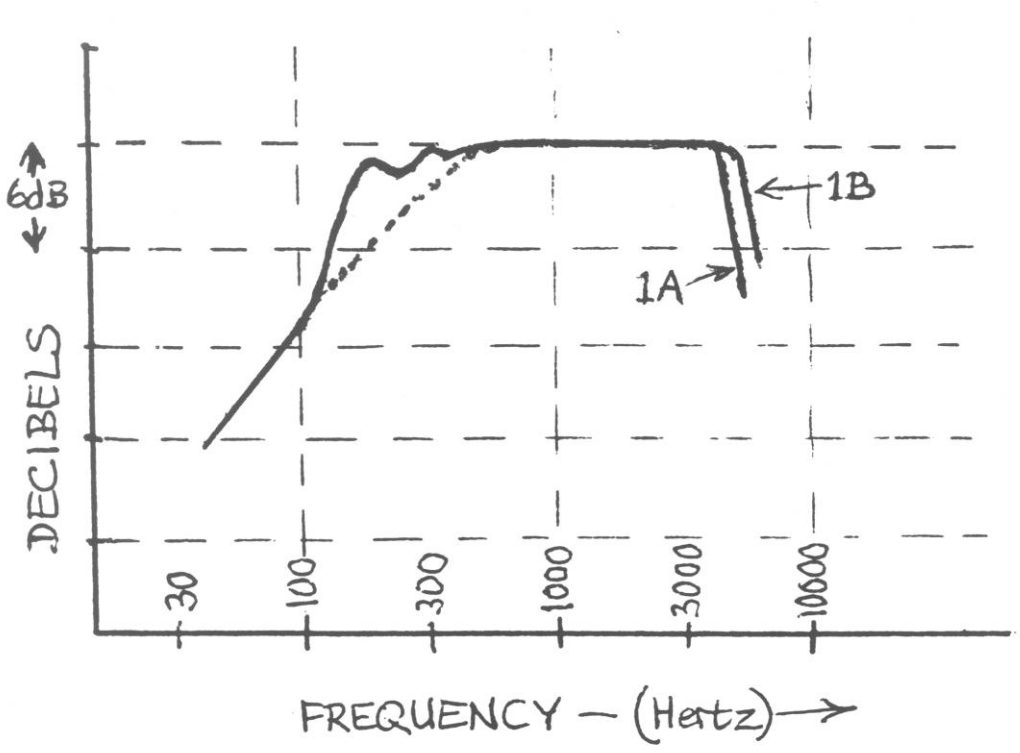


Fig. 7. Responses of Western Electric 1A and 1B cutters.

The reason this doesn't seem to have been noticed by anyone apart from the original designers was that another effect partially cancelled it. The rubber line did not always behave like an ideal "resistive element" at lower frequencies (Ref. 14). Different accounts say the same thing in different ways, but the normal bass cut due to the coil resistance was partly neutralised by a resonance at about 160Hz. Sound waves travelled along the rubber line torsionally at about 3000 centimetres per second (Refs. 13 and 15). The rubber line was about nine inches long, and when it wasn't new, it seems vibrations might be reflected back into the armature a few milliseconds later, giving a bump at 160Hz. Although it neatly filled in the bass-cut due to the coil, to reverse-engineer it today we must mix-in antiphase signals delayed by about five milliseconds with a top-cut at 12dBs per octave (this would be how they propagated through a rubber line with mass and resistance). This gets rid of the "lumpiness" audible in many such recordings.

There is an unpublished British HMV frequency test record with the matrix number (no prefix) 5729 I Δ. It comprises only fixed frequencies from 8kHz downwards, and unfortunately there isn't one between 200Hz and 130Hz. But the others depict an accurate Blumlein shape (Fig. 2) equivalent to "Blumlein 250Hz."

Continuous development kept the "Western Electric 1-something" in the forefront in America, but British record companies found the royalties burdensome, and great efforts were made to find alternatives. By 1932 alternatives were in place for all the original British record makers with one exception. This was the EMI Mobile Recording Van, because the alternative required high voltage supplies which could not easily be obtained from accumulators; but Blumlein equipment was finally installed in 1934.

6.21 How to recognise recordings made with Western Electric 1A and 1B systems

Ariel, Gramophone Co. (including HMV and foreign equivalents), Hispanophone, Regal-Zonophone, Zonophone, and Victor and Victrola repressings from the aforementioned: A triangle after the matrix number.

Ariel, Columbia, Daily Mail "Brush Up Your French", Hugophone, Regal, Regal-Zonophone: A capital W in a circle somewhere in the label surround (often in front of the matrix number), or on pressings from post-1945 metalwork, a matrix number beginning with W not in a circle.

Odeon, Parlophone, Parlophone-Odeon: A W in a circle as above, unless there is *also* a £ in a circle, in which case it was made by the Lindström system (section 6.27 below). Fortunately the same equalisation techniques apply to both, so I shall not consider the matter further; but listen out for the bass cut due to Lindström's microphone or amplifier!

Bluebird, Victor, Victrola: If repressed from a matrix "Recorded in Europe" with a triangle, the same as Gramophone Co. above. Otherwise sometimes the letters VE in an oval somewhere in the label surround, and/or the words "Orthophonic Recording" (not "*New* Orthophonic Recording") on the label.

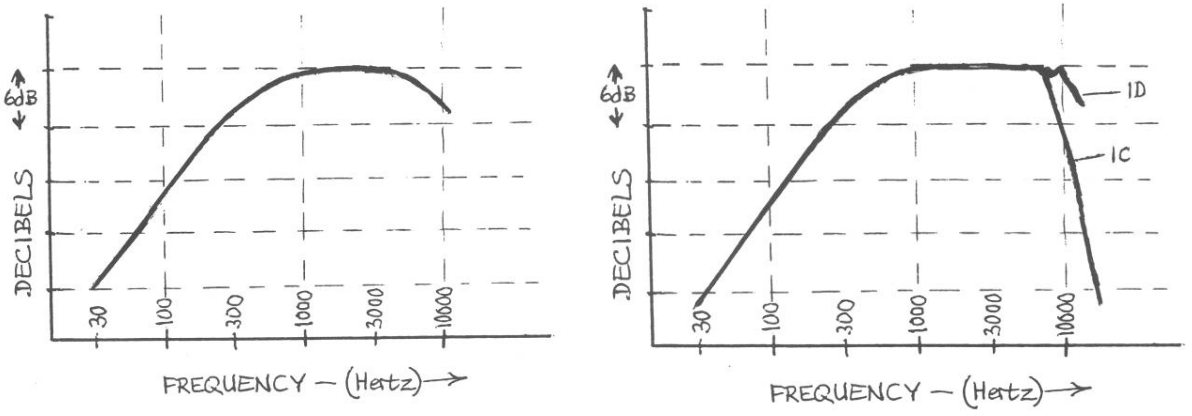
Homochord: Some records were made by the British Gramophone Company for the Homochord logo using this system in the years 1926 to 1928. They can be recognised by matrix numbers beginning with the prefixes HH, JJ, HR, or JR. Homo Baby, Sterno Baby, Conquest, Dixy, Jolly Boys. These are all logos of six-inch records made by the British Gramophone Company for Homochord in the year 1926; they all have matrix numbers prefixed by Ee.

6.22 Summary of equalisation techniques for the above

The above recordings should be equalised constant-amplitude below 250Hz (636 microseconds) and constant-velocity above that. When you are certain a Western Electric microphone was used (which was the case for all the big record companies pre-1931), a dip of 7dB at 2.9kHz may also be applied. You might also reduce any "tubbiness" by mixing in antiphase a portion of the signal with a 12dB/octave HF slope (above about 50Hz) and about five milliseconds delay. It might also be worth researching the settings shown on HMV Artists' Sheets, which will also give some idea of the amount of dynamic compression performed by the engineer; but currently, we do not know the exact characteristics of the bass-cut filters used after May 1927, and if we ever do, it may prove impossible to restore the full bass without contributing "rumble."

6.23 Western Electric developments after 1931

In the early 1930s, RCA Victor made several improvements to the Western Electric system. The curve in Figure 8 is taken from a sweep frequency record Victor 84522 (also known as 12-5-5). It cannot be dated exactly, but the label design suggests 1931 or 1932. This disc shows that RCA Victor had managed to extend the performance considerably, which is obvious from listening to contemporary RCA Victor records. The treble response is flat (constant-velocity) to 5kHz and droops only very slightly after that, reaching an amazing figure of only -4.5dB at 10.5kHz, the highest frequency on the disc. This suggests high frequencies were being compensated electronically, albeit with the limitations we saw in section 6.8. The bass response is smoothed out, although Moyer says this did not happen until 1938. It comprises a constant-amplitude section with a -3dB point between 500 and 600Hz (318 to 250 microseconds). I am confident this is a real performance, although it may have been cut on a souped-up machine.



Figs. 8 and 9. Responses of 1C and 1D cutters, as modified by RCA Victor.
(Fig. 8 taken from Victor 84522; Fig. 9 redrawn from Moyer Figures 4C and 4D).

Between December 1931 and February 1933, RCA recorded some "long-playing" records running at 33 1/3rpm. Although I cannot date it, a disc from RCA's "Technical Record Series" appears to document the characteristics of these long-playing records (it is pressed in similar translucent red material). Its catalogue number is 12-5-25V and its matrix number is 460625-6. Although it carries a range of fixed frequencies rather than the continuously varying tone of Victor 84522, the characteristic seems very similar. But it is clearly a very coarse groove. (It won't give a satisfactory Buchmann-Meyer image unless the groove wall is illuminated by a light less than thirty degrees above the disc surface). In the circumstances, it is difficult even to focus on the image! But it is clearly a Blumlein shape (section 6.7). It is substantially constant-velocity above 1kHz, drooping by something like three decibels at 12kHz (the highest frequency on the disc). The -3dB point between the constant-velocity and constant-amplitude sections is between the bands of 700Hz and 500Hz.

About 1931, RCA invented the "ribbon microphone," with a much more smooth and extended frequency response. However, it was found that the peak of the Western Electric microphone at 2.9kHz had been beneficial in overcoming surface noise, so it seems an electronic treble lift starting at 2.5kHz (63 microseconds) was added to emulate this, although it isn't on Victor 84522. However, Moyer (Ref. 11) is not very clear about the chronology; according to him, it wasn't until 1938 that the entire chain from microphone to pickup was changed. (He makes other mistakes about chronology). Whenever it may have been introduced, I call the new system "Western Electric 1C." It comprised a cutterhead modified by RCA Victor, with a smoothed-out bass response and deliberate equalisation to compensate. It was used with "improved recording channels", which I assume is Mr. Moyer's expression for new electronic equipment. He says the pre-emphasis found to be advantageous with ribbon-mikes was moved into the "recording bus" (so it was permanently in circuit, giving constant-amplitude above about 2.5kHz or 63 microseconds), the adjustable bass filter was discarded, and a low-pass brick-wall filter was imposed at 8500Hz "primarily to reduce noise and distortion effects resulting from playback turntable flutter, pickup tracking, and manufacturing methods." (*sic*)

We know that by 1943 RCA were selling their own wide-range cutterhead; but Moyer is adamant that Western Electric equipment, still working on the same principles, was upgraded yet again "about 1947"; I have called this the Western Electric 1D. He suggests that the 1D was essentially the same as the 1C without the brick-wall filter.

Moyer points out that when standardisation of all RCA Victor records was attempted in 1938, a 6dB/octave slope below 500Hz (318 microseconds) was usually found to be satisfactory for most older records. This writer agrees for post-1932 RCA Victor records, and the assertion is almost identical to the evidence of Victor 84522; but it is probably an oversimplification for earlier Victors.

Because I do not live in America, I cannot establish whether similar curves were used by other manufacturers. My only such (aural) experiment compared an early Columbia LP, assumed to have the characteristic defined in section 6.68 below, with its 78rpm equivalent. The 78 proved to have a bass-cut at 500Hz (318 microseconds). And purely aural evidence (*not* side-by-side comparisons) suggests a similar curve was used by the ARC/Brunswick companies which evolved into US Decca.

6.24 Recognising recordings made on RCA-modified and Western Electric 1C and 1D systems

Bluebird, RCA Victor, Victrola: all records mastered in the New World from 1932 until 1949; Gramophone Co. (including HMV and foreign equivalents, and a few Regal-Zonophones): A diamond after the matrix number (◊); matrixes with the prefixes A or 0A are European reissues of American material. There wasn't a consistent policy of dubbing them to new matrixes. Sometimes the pre-emphasis was removed to bring the records to European standards (because clockwork gramophones were still popular), and sometimes the dubbings were made without the pre-emphasis being removed, possibly to maintain consistency in multi-record sets. It is not easy to describe the different appearance of such a record, since British records from both original matrixes and from dubbings carry the same matrix number. It is a case where experience counts; but one indication is the eccentric run-out groove. From 1933 it is "single groove" on all British dubbings, but the "double groove" of the original American metalwork continued until March 1940 (at least as far as matrix number 0A 048176◊).
Columbia, Parlophone, and Regal-Zonophone reissues from US Columbia and Okeh sources: A capital W in a circle somewhere in the label surround (often in front of the matrix number), or a matrix number beginning with W not in a circle.
Western Electric 1C and 1D systems were used for many other electrical records on the American continent without any identification.

6.25 Summary of equalisation techniques for the above

The above recordings should be equalised constant-amplitude below about 500Hz (318 microseconds) or perhaps higher, and constant-velocity above that. I have not had enough experience to say definitely when 2.5kHz de-emphasis (63 microseconds) may be found. As the technique seems not to have been publicised at the time, I haven't called this an "official" Recording Characteristic; but the story will continue when we consider such characteristics in section 6.70.

6.26 Other systems using line-damped cutterheads: British Brunswick, Decca, DGG 1925-1939

Victor's smaller competitors didn't wait long to get their hands on the new technology. By May 1925 the Brunswick-Balke-Callender company had electrical recording in its Chicago studio, and aural comparisons show suspicious similarities to the sound of Victor's recordings. Because both parties were deliberately being secretive, it is impossible to know exactly what went on today; but I suspect the secrets went with W. Sinkler Darby, a Victor expert who defected to Brunswick about this time.

Brunswick may have erected a smoke-screen to conceal what they were up to. They claimed to use the General Electric Company's "light ray" system of recording. G.E.C. took out three patents on the subject. One deals with what we now call a "microphone", which comprised a tiny mirror vibrated by sound waves reflecting light into a photo-electric cell. (Ref. 16). This was similar to the principle of the "optical lever" used in scientific laboratories for measuring tiny physical phenomena. This microphone

could well have been used, although it was subsequently shown that any extended high-frequency response would be compromised by lack of sensitivity. I am very grateful to Sean Davies for the suggestion this may have been the microphone used by the Deutsche Grammophon Gesellschaft until about 1930. Certainly DGG recorded items which became issued outside Germany on the Brunswick label; they are characterised by an unequalised diaphragm resonance, like the sound made by an undamped telephone earpiece of the time. We do not yet have technology for compensating this.

General Electric's other two patents, for a moving-iron cutterhead built round a strip of steel under high tension so its resonance would be ultrasonic (Ref. 17), and an active mixer for combining "light ray" microphones for greater output (Ref. 18), simply would never have worked as the patents showed. (Indeed, the latter was disallowed by the Patent Comptroller in Britain). What, then, did Brunswick actually use?

History is silent until 1932. In the meantime, Brunswick set up a recording-studio in London (with audibly better microphones), the company went bankrupt, and the studio fell into the hands of the newly-formed Decca Record Company. On 1st July 1932 Decca patented a new kind of swarf-pipe for removing the shavings from the wax (Ref. 19). The patent drawing depicts something looking identical to a Western Electric cutter (although significantly the damping wasn't shown)! And some of Decca's engineers later recalled that, although they were concealed in home-made cases, they were "Chinese copies" of Western Electric rubber line recorders, manufactured by Jenks & Ader in the USA.

6.27 Other systems using line-damped cutterheads: The Lindström system

To identify the records upon which royalties were payable to Western Electric, we saw earlier that legitimate manufacturers marked their stampers. The Victor/Gramophone companies used a triangle, while Columbia and the German Lindström organisation used a W in a circle. In 1928 Lindström started marking their matrixes with a £ sign in a circle instead of a W in a circle. It appeared on many Odeon records (Parlophone in the British Empire); and when British pressing work was taken over by the Columbia company in 1929 the pressings acquired a W as well. (This was because Columbia's Wandsworth factory was adding the W in blissful ignorance of what it meant!)

Clearly, Lindström had made their own version of the equipment. I am very grateful to Sean Davies for drawing my attention to a memo in the EMI Archives dated 28th December 1931. This shows that the Parlophone studio in Carlton Hill London had both Western Electric and Lindström systems, although the former was hardly ever used at that stage.

The earliest known recording showing the £ sign was made on 24th October 1928 (matrix number 2-21025, published under catalogue number P8512 (Germany) and E10839 (Britain)). The £ sign continued until about 1936 when the Blumlein system (which I shall consider in section 6.29) was adopted throughout the EMI organisation. Since not every country had access to pound-signs on their typewriters and matrix-punches, it is thought that the letters L and P serve the same function.

Some frequency test discs were made in Berlin in about 1929 which enable us to assess the frequency response of the Lindström cutter electronics and cutterhead; the matrix number of the sweep-tone side is XXRP2. They were issued in a commercially-issued set (Parlophone P9794 to P9798) with £ signs. Unfortunately it has not been possible to date them, because they are in a special matrix series; but the first three were

published in Britain in August 1929. The labels attribute them to "Dr. Meyer of the Hertz Institute". Dr. Erwin Meyer was a co-author of the paper on the Buchmann-Meyer image (Ref. 20).

The original paper includes a photograph of this record. It is uncanny how similar the performances are to a poor Western Electric cutterhead. Both resonances on XXRP2 are conspicuous to the naked eye. I speculate that it was these discs which inspired Meyer to write his paper. The resonances amount to +5dB at 5kHz for the treble resonance and +6dB at 150Hz for the bass resonance.

The other side of P9794 has a sweep "howling tone." (Nowadays we use the term "warble-tone"). I speculate that Lindström's system was modified to improve the response in view of Dr. Meyer's discoveries. It has the matrix number XXRP3 Take 2, and measurement shows the resonances are better damped (especially the low-frequency one).

Incidentally, both records claim to have tones down to 100Hz. This is not correct. The lowest frequency on both sides is 130Hz.

The series also includes widely modulated warble tones on P9796 (matrixes XXRP6 take 2 and XXRP7). Analysis of these gives frequency responses in agreement with matrix XXRP3-2. The other records all comprise fixed tones, from which we can gain no useful evidence.

Listening suggests a £ sign is sometimes associated with a very considerable lack of bass, which may be due to the microphone or its microphone amplifier. I have been listening to Rosenthal's recording of the Chopin First Piano Concerto, the first movement of which is "normal" and the other two have the extra bass cut, in an attempt to match them and estimate what was going on. The two sessions did not have the microphones in the same position, which makes rigorous comparisons impossible; but my conclusion is that one needs an extra 500Hz lift (318 microseconds) in addition to the usual 250Hz (636 microseconds), and even then the bottom octave doesn't come right.

Through the good offices of David Mason, I was able to purchase a second-hand single-sided Odeon "Frequenz Platte" (catalogue number 1528b, matrix number FQ 1528b). It is a centre-start disc – the frequency sweeps up rather than down - and if the 1kHz tone is accurate, it needs to be played at a little over 80rpm. This too carries a Blumlein-shaped frequency-sweep, except that the very extreme bass has been *boosted* (that is, constant-velocity) in the manner common to later microgroove discs. This is rather difficult to measure on my copy because it has become worn; but the overall flattest response comes with 398 microseconds and 1591 microseconds. There is also a droop in the extreme treble, which may be an effect of the cutting stylus rather then the actual modulation, and begins to take effect above 8kHz. If anyone can help, I'd like a date for this disc. I estimate it is about 1936, but I could be considerably in error.

6.28 Conclusion to line-damped systems

So, quite apart from the official Western Electric licensees, there is now evidence that several other recording companies were using something very similar. The final evidence is Courtney Bryson's book, published as late as 1935, which gives a lot of space to instructions on how to fiddle such cutterheads with files, screwdrivers, and rheostats (Ref. 21). Despite careful patent protection, Western Electric's rubber line system was being used very widely indeed, and I don't believe we have yet reached the end of the matter.

Because I do not live in America, I cannot establish whether similar curves were used by other manufacturers. My only such (aural) experiment used an early Columbia LP, assumed to have the characteristic defined in section 6.68 below, compared with its 78rpm equivalent. The 78 proved to have a bass-cut at 500Hz (318 microseconds). And purely aural evidence (*not* side-by-side comparisons) suggests a similar curve was used for 78s made by the ARC/Brunswick companies which evolved into US Decca, while *side-by side* comparisons have shown that the earliest Capitol 78s depict a bass cut at 1kHz (159 microseconds).

6.29 The Blumlein system

Alan Blumlein is now remembered for his pioneer stereo experiments, but I shall be talking about his monophonic system and its derivatives. The recording characteristics he espoused were emulated by many other European manufacturers, there is a fair amount of objective evidence, and we saw in section 6.13 that "simple" cutterheads gave this curve by default. In section 6.7 I defined the meaning of a "Blumlein-shaped" frequency response; the sections until 6.46 will be describing these. I shall start by describing Blumlein's system (used for commercial recording throughout most of the "Old World" by labels of the EMI Organisation), and then I shall consider other systems known to give similar characteristics.

Blumlein was recruited by English Columbia in 1929 to circumvent Western Electric royalties. His system seems first to have been used commercially on 22nd January 1931. But eight weeks later Columbia amalgamated with HMV to form EMI Ltd. HMV was in the process of building a new studio complex at Abbey Road where Western Electric gear was to be installed, and recording started there the following September. Between December 1931 and July 1932 the systems were compared. Blumlein's won, and gradually all EMI's recording facilities in Britain and overseas were converted.

The novel feature, which enabled Blumlein to dodge the Western Electric patents, was that resonances were not damped mechanically. Instead they were neutralised by a combination of eddy-current damping and electronic equalisation. This applies both to the microphones and the cutterheads, so the patents are interlinked.

Equalising resonances with antiresonant circuitry is rather like fighting fire with fire. If things are not exactly right, matters easily get out of control. It is reliably reported that Blumlein (and his mechanical engineer Herbert Holman) pressed their assistants to calculate the moments of inertia of complicated metal shapes theoretically, and chastised them if the resonant frequencies of the manufactured prototypes were not correct within one percent. But this sort of precision was vital if peaks and troughs in the frequency response were to be avoided. After the circuitry was tuned, the results should have stayed consistent because there was no rubber or other perishable materials; but peaks and troughs are exactly what make Blumlein recordings so distinctive - to my ears, anyway!

6.30 The Blumlein microphone

The Holman-Blumlein microphone (known as the HB1) was a moving-coil type with a diaphragm made of balsa-wood coated with aluminium foil. British patent 350998 refers to a primary resonance "at or below 500 cycles per second." The original design had a

field magnet, but by 1934 a version with a permanent magnet (known as the HB1E) was also used. Examples of each are preserved at the EMI Archive.

Both mikes had the same front-dimensions, but the HB1E was less deep. They had chamfered edges and slightly smaller diameters than Western Electric mikes, so the axial pressure-doubling was less pronounced. But their bulk still meant high frequencies from the sides and rear were attenuated.

Frequency curves are given in Refs. 22 and 23. Ref. 22 does not name the microphone, but shows a picture of an HB1E. Ref. 23 actually says it is an HB1E and adds a polar diagram. The on-axis curves show a 5dB rise from 2 to 5kHz, presumably the pressure-doubling effect, and the latter reference shows the output remaining above the 1kHz level up to the limits of the graph at 15kHz, albeit with some wobbles. It was claimed "the rigid structure of the diaphragm causes the first nodal resonance to occur at 17 kc/s, below which it moves as a single resonant system"; but neither the graph nor this writer's listening experiences support this, so maybe further research is needed.

It could be argued that the shelf in the high-frequency response outweighed the mike's insensitivity to HF from the sides and back. This would be reasonable for musical sources of finite size (rather than point sources), especially in a reverberant environment; and the mike was renowned for its performance on pianos (Ref. 24) so the graphs actually support this theory. I do not think we need to address the frequency response aberrations - on piano recordings, anyway! As it lacked the cavity resonance of its Western Electric predecessors, and its primary resonance was equalised by a circuit included within the microphone cable (shown in the patent), current practice is to consider it "flat."

By the end of the second world war the HB1E microphone had given way to other types, including an EMI ribbon mike (also shown in Ref. 23), and early Neumann condenser mikes. Again, current practice is to consider these "flat."

6.31 The Blumlein moving-coil cutterhead

A passive equaliser (that is, one having no amplification) was placed between the power amplifier and the cutterhead to correct resonances and give the required characteristic. The cutterhead used a novel principle (described in British Patent No. 350954) which made it virtually overload-proof; but the equaliser was described in Patent No. 350998. It had no less than 12 elements, and presumably shows the state of the art just before the patent applications were lodged.

Each individual cutterhead had its own equaliser, and there were strict instructions to prevent cutterheads and equalisers being swapped. But there were at least four subsequent versions of the cutterhead, and the equaliser components were designed to be easily changed by maintenance staff (not everyday operators). Since we do not know which equaliser and cutterhead was used for which record, further refinements do not seem possible today; but EMI engineering documentation always shows "Blumlein shape" was the target.

We saw another snag in section 6.8. To pull some curves straight by electronic equalisation implies infinite amplification, difficult with limited power to spare! In practice, both microphone and cutterhead systems "drooped" at the extremes of the frequency range.

6.32 Test discs made by Blumlein's system

I shall now describe three frequency records made by Blumlein cutters. The earliest is TC17, a record made in the spring of 1933 with matrix number (no prefix) 5403 □ I. It was manufactured in large quantities for testing HMV radiograms. It starts with a brief frequency sweep, and the same groove continues with dubbings from records which were known empirically to give reproduction problems, such as cabinet rattles and motor-drag. Thus it probably shows the actual performance of a Blumlein recorder in 1933, rather than an idealised laboratory performance. The sweep runs from 8kHz to 30Hz at "normal recording characteristic." Although it generally follows a "Blumlein 200Hz" shape (795 microseconds), it is by no means perfect. Above 2kHz there is a gentle roll-off in the treble, amounting to -1dB at 5kHz and -4dB at 8kHz, with slight ripples along the way. Below the turnover frequency at 200Hz there are signs of a massive but ill-tuned resonance; the response swings from +3.5dB at 180Hz to 0dB at 170Hz and +1.5dB at 160Hz in a manner reminiscent of FM demodulators. Its presence so near the turnover frequency makes it difficult to determine the intended characteristic accurately.

The next disc is numbered TC20, "Test Record for Electrical Pickups" (matrix number 2B4977 Take 5, mastered in No. 4 Studio, 19th April 1934). There had evidently been trouble with pickup armatures sticking to polepieces, because the record includes a band at 150Hz "for Freeze-over Test" which is about ten decibels louder than the rest of the record, which has eleven other fixed frequencies. The label doesn't claim any particular characteristic, but the treble droops in a similar manner to TC17. The bass is considerably lifted; but I should explain that in those days the bass lift to equalise a "Blumlein shape" was hardly ever achieved electronically, but by the pickup arm resonating on the armature compliance at about 150Hz. I therefore suspect this disc was primarily used for checking such pickups weren't thrown out of the groove, and it may not be meant as an accurate "Blumlein shape" at all.

In section 6.11 I mentioned the published disc HMV DB4037, cut in 1936. This shows no significant treble droop at the highest frequency (8.5kHz). It changes to constant-amplitude at 500Hz (318 microseconds), while the fixed-frequency tones on other discs of the set (DB4034-7) suggest a turnover at 150Hz (1061 microseconds), and the leaflet implies 250Hz (636 microseconds). It is too early to say whether these differences were consistent in any way. The result is that all Blumlein recordings made between 1931 and about 1944 are currently equalised constant-amplitude below 300Hz and constant-velocity above that. Occasionally there is strong subjective evidence that these transitions aren't correct, but when I'm not sure I use 300Hz, because it's roughly the geometrical average.

All these test discs have frequencies up to at least 8kHz (sometimes drooping slightly), but Blumlein's system was not allowed to go as high as this in practice. There is an EMI memo written by B. Mittell on 27th October 1943 (when the idea of extending the range was going ahead), entitled "Experiments in Semi-High Quality Recording." He lists experimental recordings which had been made between 28th March and 12th April 1935 "with equipment which cut off between 7,000 and 8,000 cycles." His memo continues: "the results were not good. . . because of unsatisfactory reproduction at the higher frequencies. The present light-weight pickup was not then in existence, nor were the record surfaces considered good enough at that time to justify recording in that region." Although the three test discs show 8kHz was quite feasible, the range was restricted to about 6kHz in practice.

6.33 How to recognise Blumlein cutters on commercial records, 1931-1944

HMV, Zonophone, other logos of The Gramophone Company, and repressings by RCA Victor: A square following the matrix number. A post-1945 matrix with no geometrical symbol is sometimes a dubbing of a 1931-1944 matrix.
Chappell, Columbia, Hugophone, Odeon, Parlophone, Pathé, Regal-Zonophone, Voice of the Stars: A C in a circle before the matrix number in the years 1931 to 1944. A post-1945 matrix with a plain C before the matrix number occasionally means a dubbing of a 1931-1944 matrix.
NOTE. A few Lindström recordings (issued on the Odeon and Parlophone labels) recorded in Strasbourg Cathedral in 1928-9, with matrix numbers prefixed XXP, have a C in a circle in the label-surround. These are *not* Blumlein recordings.

6.34 Summary of how to equalise the above

Equalise to a "Blumlein shape." Start with Blumlein 300Hz (531 microseconds), and only vary it if you feel a compelling need to. For serious work, log the turnover you have used. Before 1936, a gentle treble *lift* amounting to +4dB at 8kHz may be added if the surface-noise permits.

6.35 The Gramophone Company system

A Gramophone Company report dated 12th June 1930 describes how HMV engineers hoped to evade Western Electric patents. Their Dr. Dutton made a trivial modification to a Western Electric microphone capsule (cutting "slots" instead of "grooves" in the backplate) in order to evade one patent; and a very conventional valve amplifier was built to replace the Western Electric push-pull design. A cutterhead invented by Angier, Gauss and Pratt (British Patent 372870) uses ideas voiced in that report.

When trials began in Abbey Road Studio 3 on 16th December 1931, the matrix numbers were followed by a "swastika," actually the mirror-image of the swastika later used by the German Nazi party. When EMI was asked for an explanation of this in 1964, the enquirer was told that it meant "recorded by the Gramophone Company system," which seems quite unambiguous; but I must also report an alternative explanation (Ref. 26). This says the mark identified a moving-coil cutter (presumably Blumlein's), and that the swastika was changed after complaints from various continental export markets.

I have not been able to distinguish between these two explanations, and it is even possible *both* are true. If the first is correct, then only the cutterhead was novel. It was in fact a classical moving-iron cutter, and one of the features (in Claims 1 to 4 of the patent) was that the recording characteristic was defined by the electrical features of the coil, as we saw in section 6.13. Fortunately EMI did not pursue these claims, or a great deal of sound recording would have been strangled at birth! However, previous systems (like Western Electric) had used the same principle but not claimed it in a patent.

The resonant frequency of the armature was given as 5400Hz in the patent, with a peak of +4dB or +6dB, depending which graph you look at. Spectral analyses of surviving discs suggest a lower frequency and a lower peak than that, although the patent shows

five designs of armature (which could have had different resonant frequencies), and the effect of the cutting stylus was not quantified.

The constant-amplitude turnover was described with the words "for example, 250 cycles per second."

The performance of the microphone and amplifier do not seem to be documented, and we only have the historic photograph of Elgar and Menuhin at their recording of Elgar's Violin Concerto in July 1932. This shows a Blumlein HB1 microphone, although the records all have "swastikas." It is not generally possible to hear any difference between "swastika" recordings and others (possibly because they used the same microphones). Current practice is therefore to treat the former as if they were recorded to a "Blumlein 250Hz" curve.

6.36 How to recognise the Gramophone Company system on commercial records

HMV, Zonophone, RCA Victor re-pressings: A mirror-image swastika after the matrix number.

6.37 Summary of how to equalise "Gramophone system" recordings

Use "Blumlein 250Hz" (636 microseconds).

6.38 Extended-Range Blumlein recordings (1943-5) and later systems

In 1943 Blumlein's cutterhead seems to have been given an extended frequency range. According to Ref. 27, 12kHz was reached, partly by replacing the original armature pivots with one steel torsion wire and one conical rubber pivot. Listening tests suggest that a primary resonance at about 500Hz is not always perfectly tuned - there is sometimes audible colouration in this region - but there is no objective evidence. Perhaps future spectral analysis methods will confirm this; in the meantime, I ignore the problem.

The "ER System" was first used towards the end of October 1943, but it was kept secret until March 1945, when the decision was made to modify the symbols on the matrixes so extended range records could be used for public demonstrations. For HMV the square had an additional diagonal line; for Columbia and Parlophone the C-in-a-circle had an additional diagonal line.

In the meantime another system was adopted, although it seems EMI engineers didn't admit it, because it came from RCA in America! Since there were no patent implications, there was no geometrical symbol, although Blumlein ER was retained in Studio No. 1 until Sir Thomas Beecham had completed some multi-sided sets.

After this the technical standard of EMI's work is such that I cannot distinguish between different recording systems. The 78s all have an extended and apparently resonant-free frequency range of "Blumlein shape" until the introduction of International Standards in 1953. Test disc EMI JG.449 (mastered 5th-6th July 1948) documents this, showing a perfect response to 20kHz with the low-frequency turnover at 250Hz.

6.39 Summary of how to equalise "Extended-Range" and subsequent systems

Use "Blumlein 250Hz" (636 microseconds) for EMI coarsegroove discs mastered between November 1943 and (as we shall see) July 1953.

6.40 Early EMI long-playing and 45 r.p.m. records

Tape was often used for mastering purposes from 1948 onwards, but a great many commercial 78rpm releases were made from direct-cut discs recorded in parallel. From about 1951 tape-to-disc transfers became normal for 78s (they can usually be identified by a letter after the take-number).

Longer-playing records such as LPs and EPs became possible at the same time, and short items might appear on EPs and long-playing albums besides 78 and 45rpm "singles." Historians of the gramophone have criticised EMI for being sluggish with the new media. Yet microgroove masters were cut as early as July 1949, evidently for export; British readers will appreciate the joke when I say LP masters were for US Columbia and 45rpm masters for RCA Victor. They had Blumlein characteristics (in this case, "Blumlein 500Hz" - 318 microseconds), even though they were not intended for clockwork gramophones! I have checked this by comparing 78s with early LPs and 45s of the same performances, assuming the former were "Blumlein 250Hz." Some early microgroove discs had properties reminiscent of 1930s 78s, with colouration in the mid-HF and treble above about 6kHz attenuated; but these difficulties were solved within a year or so.

It is always easier to say when something started than when it stopped, and the same goes for this characteristic. I had to compare dozens of 78s with LPs and 45s of the same performances, after which I could say with some confidence that all three changed at about the same time. The critical date was 17th July 1953 or a few days later (Ref. 28), after which comparisons are consistent with what later became the two 1955 International Standards (section 6.7).

Unhappily matrix numbers cannot give foolproof boundaries, because EMI sometimes had twenty takes before cutting a perfect LP master (although the situation wasn't quite so bad for 45s), and HMV's 0XEA and 2XEA numbers seem to have been allocated well in advance (not in chronological order). In July 1953 Abbey Road was mastering microgroove in duplicate, and the take-numbers mean master-discs as opposed to artists' performances. If a pair of masters failed, it would take three to eight weeks to get them cut again. Most previously issued Blumlein versions were later remastered to conform to International Standards, sometimes getting new matrix numbers in the process, in which case information for older versions has now disappeared. And I have found at least one case where the new version was still Blumlein 500Hz, presumably because of what was on the other side of the record; evidently changes in sound quality were avoided.

I hope the following table will help you identify the last Blumlein versions by matrix number, but please remember most must be approximate, and can only apply to Take Ones and Twos.

HIS MASTER'S VOICE

LPs: (10-inch and 12-inch) Between 2XEA213 and 2XEA392, and at 0XAV145.
EPs: 7TEA 19, 7TAV 28
SPs: Between 7XBA14 and 7XBA21, and at 7XCS 23, 7XLA 2, 7XRA 30, 7XSB 6, 7XVH 70
SPs: 7XEA688, 7XAV227 (Both series then jump to 1000)
78s: Between 2EA17501 and 0EA17576

COLUMBIA

LPs: (10-inch and 12-inch) Between XA561 and XAX817; XRX12
EPs: 7TCA 7, 7TCO 6
SPs: 7XCA185, 7XCO 87 (Both series then jump to 1000)
78s: Between CA22600 and CA22610, and at CAX11932

PARLOPHONE

LPs: XEX 60
EPs: Probably all International Standard
SPs: 7XCE135 (Series then jumps to 1000)
78s: Between CE14643 and CE14689

MGM

My comparison method falls down for most MGM records mastered in America, because I haven't anything else to compare them with. The following are British.

SPs: 7XSM203 (Series then jumps to 1000)
78s: 0SM420

REGAL-ZONOPHONE

78s: CAR6800

6.41 Other systems giving a "Blumlein-shaped" curve - amateur and semi-pro machines

I shall now list other systems known to have a characteristic comprising a constant-amplitude section below roughly 300Hz and a constant-velocity section above that. Virtually all disc-cutting machines meant for amateur or semi-professional use followed the principles outlined in section 6.13 which gave these results by default, unless altered by electronic means. You may find such discs (both coarsegroove and microgroove) in the form of "acetates," "gelatines," or short-run pressings.

In Britain, M.S.S. sold disc-cutters which dominated this market, and its director Cecil Watts maintained close co-operation with the BBC. A couple of pre-war BBC frequency test pressings exist mastered on M.S.S machines, which document Blumlein 300Hz objectively. During the war the strategic importance of M.S.S was such that it was taken over by the British Post Office, and development of the cutter was placed on a slightly more scientific footing; details are given in Ref. 29.

M.S.S cutterheads were comparatively cheap, and appealed to the man who could build his own amplifier and mechanism. A "Blumlein shape" will nearly always be appropriate in these cases. But M.S.S also sold its own electronic equipment and mechanisms, the former including the "Type FC/1 Recording Characteristic Control Unit" (1954) giving: "British" (i.e. Blumlein 250Hz); the BBC's characteristic; and the "N.A.B. Lateral" characteristic (which will feature in section 6.65, but by 1956 it had been superseded). There was also circuitry for extending the cutterhead's frequency range and flattening its resonance, and for compensating the end-of-side losses at the middle of the disc. In a well managed studio, these could have given almost professional results.

I was brought up with post-1956 M.S.S cutterheads, and can confirm that the fundamental resonance was 7kHz and the bass roll-off due to the coil was at 500Hz (318 microseconds). When International Standards matured, this didn't give the 75 microsecond pre-emphasis needed for RIAA; a graph supplied by the company showed the additional equalisation needed, but not how to achieve it. (Ref. 30).

I should like to give you the equivalent figures for cutterheads supplied on other semi-pro disc-lathes, such as B.S.R., Connoisseur, EMI, Presto, and Simon; but in the case of the EMI machines, and sometimes the others, the recording amplifier included "tone controls" for subjective use, so the exact characteristic cannot be defined anyway. Many M.S.S and Presto machines were subsequently fitted with a Grampian cutterhead, a commercial version of the BBC Type B which could be used with non-motional negative feedback. (This would neutralise some or all of the bass-cut due to the coil). The resonance was 10kHz, thereby reducing S-blasting in the presence of pre-emphasis; but it was not available until 1953.

6.42 British Decca-group recordings 1935-1944

In section 6.26, I described how the early days of Decca seem to have been founded on a clone of a Western Electric cutter. The company's chairman Edward Lewis thereupon took over most of the bankrupt record companies in London, acquiring several original pieces of recording equipment in the process, and it would be extraordinary if the best of each were not adopted. In the summer of 1935 Decca published frequency test disc EXP55 (later issued in Australia on Decca Z718). Although it comprises only fixed frequencies, it documents "Blumlein 300Hz" up to 8kHz.

It was not until mid-1937 that Decca took over the Crystalate group, which itself had been collecting equipment from other sources. Crystalate had recruited Arthur Haddy in 1931 to improve the (American) Scranton Button Co. equipment (which had also been supplied to the UK Dominion Record Company, see section 6.14). In a 1983 interview (Ref. 31), Haddy mentioned several subsequent pieces of cutting equipment, starting with a "Blue Spot" loudspeaker unit modified by himself, a cutter from Eugene Beyer of the German Kristall company, the first Neumann cutters, and a clone of the Western Electric cutter supplied by Jenks & Ader of America (in use when Decca took over Crystalate). In that year (1937) Haddy's own moving-coil cutters were adopted throughout the group, since Haddy considered this was the only way to avoid S-blasting when stretching the performance of moving-iron cutters. These evolved in secrecy for the next twenty years, because the resonance was damped by the same means that Blumlein had described in British Patent 350998. Meanwhile, test disc EXP55 remained in the catalogue for another ten years, which I regard as evidence that there was no change of equalisation policy throughout all this.

In 1944 Haddy's famous "full frequency range" system was implemented. As this used a pre-determined "non-Blumlein" characteristic, I shall discuss it in Section 6.66; but I believe "Blumlein 300Hz" will be appropriate for matrixes until number DR8485-2.

6.43 Summary of how to equalise Decca-group recordings 1935-1944

Use "Blumlein 300Hz" (531 microseconds)

6.44 Synchrophone

This was a company formed in 1931 to make use of a record factory in Hertford, England. The factory had changed hands several times during the years, and was vacant again after the collapse of the Metropole Record Company (makers of Piccadilly and Melba records). It is thought (but it's not absolutely certain) that the recording machinery passed from Metropole to Synchrophone; at any rate, the matrix numbers are contiguous.

Synchrophone specialised in making records which would not have the usual legal complications when they were publicly performed. They were especially targeted at cinemas for providing interval music, when they have the logo "Octacros." Some of these were reissues of the "Piccadilly" label, but new recordings for the Octacros label followed.

There is no known written evidence about the recording machinery, but various frequency test discs survive which provide *embarras de richesse*. I shall cut a long story short and tell you about the set of frequency records which I personally consider you should *ignore* - a set of ten single-sided 78rpm engineering test records, presumably for checking cinema sound systems, numbered from Tech.90 to Tech.99. They are so appallingly bad that it is difficult to see why they were published. But there is one frequency record which seems to make more sense. Several test pressings with the matrix prefix ST survive. One, ST.29, carries two frequency runs in about two minutes of groove, one comprising a sweep, and one fixed tones. It looks to me like a straightforward measurement of the performance of the cutter, rather than a disc intended to be used in anger; but its performance is much better than the series mentioned above. Better still, it seems to document the right curve for reproducing contemporary recordings – "Blumlein 300Hz."

6.45 Summary of how to equalise Synchrophone recordings

Use "Blumlein 300Hz" (531 microseconds).

6.46 Conclusion to "Blumlein-shaped" equalisation

I am quite certain other manufacturers used "Blumlein characteristics"; unfortunately, I have no objective evidence. For example, I am sure Deutsche Grammophon used "Blumlein shape", extending their frequency range at about the same time as EMI in Britain. Coarsegroove Polydor test discs were marketed at the end of the second World War; does any reader have access to one?

6.47 The Marconi system

There is a definite demand for the information in this section; but it isn't complete, and several problems remain. So I am writing this in the hope that further research may nail things down.

The British Marconi company was probably the best equipped of all British companies to develop a new electrical recording system, and it was advertised as being exclusive to Vocalion. The first published matrix seems to be dated "August 11, 1926" (Ref. 32); and the system continued to be used until Vocalion's studio was closed down after its takeover by Crystalate in 1932, that is to say about the end of 1933.

Many of its records have an M in a circle printed somewhere on the label. The following logos are known to have used the system at some time in their lives. Sometimes only a handful were electric; but you can assume that if it's an electric record of British origin on any of these labels, it must be the Marconi company's process: Aco, Aeolian-Vocalion, Broadcast, Broadcast Four-Tune (I believe), Broadcast Junior, Broadcast Twelve, Broadcast Super-Twelve, Coliseum, Scala and Unison. Also some Beltona issues (between catalogue numbers 1194 and 1282 - there was a change to Edison-Bell after that); many Linguaphones; a few Vocalions up to catalogue-numbers X10029 A.0269 and K05312 (later British issues were made by Decca from US metalwork); and National Gramophonic Society (catalogue "numbers" HHH to TTT and NGS.65 to NGS.102). Finally, some of the Broadcast Super-Twelves were reissued from original metals on the Rex label. Most early Rex records reproduce the catalogue number in the label-surround, and if there is a second number in a 3xxx series that is the Broadcast Super-Twelve *catalogue* number, and the Marconi process can be assumed.

After 1934 it is thought the recording equipment was sold to Linguaphone, who continued to use it until at least the second world war for recording new language lessons. It is also known that the system was used by the equivalent Vocalion company in Australia until about 1934.

I have discovered no written information about the technical aspects, except that it was a moving-iron system (Ref. 31) Listening tests suggest that at least three kinds of microphone may have been used at different times, and the first couple of years' recordings suggest that they were cut by a "low resistance" cutterhead, since constant-velocity seems to give the correct musical balance above about 500Hz. The Marconi Company developed the Round-Sykes "meatsafe" microphone, which was also used by the BBC. A frequency curve for this microphone is given in Ref. 34, and the somewhat tubby bass of the earliest discs is consistent with that curve; but I do not consider we can equalise it with rigour, because there was a great deal of empirical adjustment. Controversy raged over whether it was better to hold the coil in place with three pieces of cotton-wool smeared with Vaseline, or four! (Ref. 35)

However, aural evidence suggests that there were two further developments at least. In late 1928 some of the early "Broadcast Twelves" were sounding very shrill and metallic, which might be a different microphone - possibly the Marconi-Reisz described in Ref. 36. But despite the lack of written information, a quirk of fate enables us to get some objective measurements of the *cutter* in the 1930s.

Two records were issued on the "Broadcast" label which we can analyse to obtain parts of the frequency response. One is Tommy Handley's sketch "Wireless Oscillations and Aggravations" recorded in about November 1931 and issued on Broadcast Super-Twelve 3133. The other is "Sandy At The North Pole," a comedy sketch recorded by Sandy Powell in about December 1932 and published on Broadcast 926 (a nine-inch

diameter disc). The plot of the latter involves his sending a radio message home. (This was the origin of his catchphrase "Can you hear me, Mother?") On both these records, it seems the Vocalion engineers plugged a beat-frequency audio oscillator into the mixing desk, and twiddled the frequency arbitrarily to simulate the heterodyne whistles common on A.M radio reception in those days. (There's no *proof* this happened, but the measured results admit no other explanation). Although there is speech on top of the oscillations, and there are places where the volume alters while the frequency stays the same, it is possible to take slices out of the recordings between the words and analyse them.

These show that (in the years 1931-2, anyway) the electronics and cutter gave a constant-velocity characteristic between 1490Hz and 4300Hz accurate to one decibel, and between 800Hz and 4600Hz within two decibels. The frequency goes as high as 5200Hz on "Sandy at the North Pole", and isn't more than 4.5dB down there; but I'm not quite certain the volume was stable at the start of this sweep. Also the nine-inch disc shows frequencies around 750Hz to be about +3dB higher. The ten-inch doesn't have this effect, and I can't explain it. However, the consistent results between 1500Hz to 3000Hz also support the theory it was a low-resistance cutter.

I have also been trying to find examples of the same material mastered at different times, so I can compare "like with like" using aural methods. So far, my best experiment uses John Gielgud's performance from "Hamlet" (Act 2, Scene 2), which he recorded for Linguaphone twice (once by Marconi's system, and once by Decca's ffrr system on 78). I am very grateful to Eddie Shaw for telling me the former was published in February 1931, and the latter in August 1955. Gielgud also recorded it a third time on LP (HMV ALP1483). The three performances are almost wilfully different, so comparisons are not easy; but there are a couple of sentences delivered at similar pitches and rhythms which permit qualitative comparisons.

The 1931 Linguaphone is noticeably brighter in the upper harmonics of the vocal cords than either of the others when played with a Blumlein-shaped curve. There are two possible explanations. One is that Marconi's system emulated Edison-Bell's (section 6.14), by equalising the mike to give a 3db-per-octave slope (but this didn't apply to the oscillator). The other is that 1931 mikes were even bulkier than Western Electric ones, so they reflected more sound back at the performers, doubling the electrical output above about 2kHz.

I find the three discs match best when I set the 1931 one to the BBC's "2dB-per-octave" curve (section 6.57). It still isn't quite right to my ear; but listening is now being overwhelmed by differences between the performances.

That's the current state of the art so far as the Marconi system is concerned. Needless to say, I should be interested if readers can come up with anything more.

6.48 BBC Disc Record equalisation - Introduction

The British Broadcasting Corporation did not always follow international (or *de facto*) standards with its discs. The next fourteen sections summarise the electrical characteristics of such discs so they may be reproduced correctly. There may be some urgency for this, because the BBC sometimes gave performers a cellulose nitrate disc instead of a fee, and such "nitrates" are not only unique, but have a limited shelf-life.

As you might expect from a large and cumbersome organisation like the BBC, there were several changes of disc recording policy over the years, but hardly ever any clean breaks. There were long periods of stability, so for 90 percent of BBC discs I can

refer you to the summary in section 6.61 below. But the other ten percent are in "grey areas," sometimes with very unexpected characteristics. I must recommend you to become familiar with the BBC's disc recording history (Ref. 38) to assess what technique was used for the particular disc you hold in your hand.

From about 1955 onwards, the BBC tended to copy its older formats onto new media. The most important examples concern Philips-Miller film recordings, copied to microgroove LPs; but as the original films have now been destroyed, we must use the LP discs. Other items were copied from 78rpm and 60rpm discs, but there is evidence that occasionally engineers of the time did not get the equalisation right, or else they were copied before International Standards were adopted. Ideally, you should have access to the "originals," whatever they may be; but if you haven't, I can only ask you to be aware of the evolution of BBC practices, so you may detect when an equalisation error has been committed, and reverse-engineer it.

I have been doing industrial archaeology research to establish what was going on. There are three basic sources of information:

(1) Surviving engineering test discs

(2) Analysis of circuitry and hardware in contemporary engineering manuals

(3) Spectral comparisons of the same sounds recorded by different equipment

I will try and indicate when one system gave way to another as accurately as I can; but the dates and matrix numbers will often be approximate. The numbering-systems for both the matrixes of pressed discs, and for "current library" nitrates, should be understood for this reason. They will be considered in sections 6.51 and 6.52 below. After this we shall study the various phases of disc recording history with consistent equalisation.

6.49 The subject matter of these sections

I cover all disc records mastered inside the BBC. The range includes:

1. Pressings and nitrates for the "BBC Archives," also known as the "BBC Permanent Library."
2. Pressings for the "BBC Transcription Service" for use by broadcasters overseas, additional copies of which were sometimes pressed for (1) above.
3. Pressings for the "BBC Sound Effects Library".
4. Records whose logo, printed in green, consisted of the words "Incidental Music." (Signature tunes and the like, mass-produced for internal consumption).
5. Nitrates cut for immediate use, administered by the BBC Current Library or one of its branches. These have a lick-and-paste label with a space for a handwritten "R.P Ref. No." This means "Recorded Programme Reference Number." Such a number has a distinct structure incorporating the code for the branch library concerned, and enables them to be distinguished from the other libraries.

The BBC also mastered records for other organisations with yet more logos. Among these I can name the "London Transcription Service" and "The Joint Broadcasting Committee" (both wartime precursors of the BBC Transcription Service), the "Forces Broadcasting Service," and "A Commonwealth Feature Programme."

6.50 Pre-history of BBC disc recording

The earliest problem is simply to recognise which discs were done upon BBC equipment. Before the BBC got its first disc cutting machines in April 1934, and for some years after that, much BBC recording was carried out by commercial record companies, mainly EMI and British Homophone (makers of "Sterno" records). The former can be recognised by matrix numbers prefixed 0BBC or 0BBCR (ten-inch) or 2BBC or 2BBCR (twelve-inch). These were mastered using equipment designed by Alan Blumlein, and the "Blumlein 300Hz" characteristic is appropriate here.

The British Homophone records have a much finer groove-pitch (actually 150 lines per inch). They usually have plain matrix numbers in the shellac, but the prefix "Homo" is added on the label. This appears to have been a "high-resistance" system (section 6.14).

Anything you find with a matrix prefix plain "BBC" and the words "Record manufactured for the B.B.C. by the Decca Record Co. Ltd." is from a nitrate cut by the BBC, not a Decca master. We shall see later that "Blumlein 300Hz" (531 microseconds) is correct.

When the labels carry the words "Record manufactured for the B.B.C. by the Gramophone Co. Ltd," a pure "BBC" prefix still means a BBC master-disc; but when they have matrix prefixes 0BBCD or 2BBCD, these may be *either* Gramophone Co *or* BBC masters. As both organizations used similar characteristics, again it will be sufficient if I say "Blumlein 300Hz" (531 microseconds).

6.51 BBC matrix numbers

Any disc which was "processed" (that is, made into a metal master so copies could be pressed) was given a "matrix number," which appeared on stampers and on finished pressings because metal cannot have a written label. The first BBC-mastered matrixes had the prefix "BBC", often hand-scribed onto a "current library" nitrate after it was finished with. Number BBC1 was a ten-inch pressing made for the BBC Sound Archive (their catalogue number 588).

When the BBC became responsible for the engineering work associated with overseas radio stations during the war, the matrixes had different prefixes to indicate the size and the pressing company. (These records were the foundation of the Transcription Service).

In September 1942 the two operations were amalgamated. With true British-style compromise, the prefixes were adopted from the Transcription Service, and the number suffixes from domestic radio, which had by then reached about 8000. An extra letter was added to determine which was which, and the following prefix code evolved:

First, 16 = 16 inch diameter
12 = 12 inches
10 = 10 inches
7 = 7 inches
Second, if there at all: F = fine-groove.
Third, P = Transcription Service Processed Disc
R = Recorded Services (i.e. domestic radio or TV).
Fourth, the company which did the galvanic and pressing work, thus:

D = Decca, Raynes Park
H = British Homophone Company
M = EMI Ltd., Hayes
O = Oriole Ltd. (later CBS)
P = PR Records, Wimbledon
R = The Transcription Manufacturing and Recording Co. (C. H. Rumble), Redhill.
RR = Rediffusion, Caerphilly
S = Statetune, Leicester
W = Nimbus Records, Wyastone Leys, Monmouthshire

Fifth, matrix number, and if not "Take 1", a take number. The take number indicates the attempts at cutting a master-disc, not different performances. The numbers formed an essentially continuous sequence, the world's largest run of matrix numbers, ending at 162695 in 1991.

Sixth, -S if stereo (confined to the early days of stereo only).

For example: 16PH meant 16-inch (implied coarsegroove) disc processed by British Homophone from a Transcription Service master, and 7FRO meant 7-inch finegroove processed by Oriole from a domestic master.

6.52 BBC "current library" numbers

Apart from Broadcasting House in London, a large number of regional centres and overseas studios made disc recordings. There were literally hundreds of these at various times; the following is only a selection. To save constant communication and delays, they allocated their own serial numbers to form current library recordings. The actual numbers were originally restricted to five digits, which were thought to be enough for a current library where recordings were not kept permanently; but by 1963, Bush House had been "round the clock" several times! Duplicate-numbered tapes were liable to appear unexpectedly, so six digits became the norm. The serial numbers were prefixed by three or four letters as follows.

First letter (if there at all):

C = a copy from another recording, whose identity was supposed to be indicated in the box marked "Source" on the label and the accompanying Recording Report.
P = Master nitrate disc intended for processing, or a backup for same.

Second letter: Format (I shall only cover disc media here; a similar system applied to tapes):

D = 78rpm coarsegroove nitrate disc, or a set of such discs
F = Fine-groove nitrate disc
M = Mobile recording (usually cut on a disc-cutter in a van or car)
S = Slow-speed (33rpm) coarsegroove nitrate disc, or a set of such discs

Third and fourth letters: Studio centre allocating the number.

AB = Aberdeen
AH = Aldenham House, Hertfordshire

AM = America (usually the New York studio)
AP = Alexandra Palace
BE = Belfast
BG = Bangor, North Wales
BM = Birmingham
BS = Bristol
BT = Beirut, Lebanon
BU = Bush House, London
CF = Cardiff
EH = Edinburgh
GW = Glasgow
GY = Germany
LG = Lime Grove
LN and LO = Broadcasting House, London, and buildings nearby
LS = Leeds
MR = Manchester
NC = Newcastle
OX = 200 Oxford Street, London
PY = Plymouth
RW = Radiophonic Workshop, Maida Vale, London
SM = Southampton
SW = Swansea
WN = Wood Norton, near Evesham

These location-codes would usually give the home of the disc-cutting machine, not of the performance; the latter would be indicated in the "Source" box on the label. Thus, DBU123456 would be an original 78rpm coarsegroove nitrate cut at Bush House, and PSOX12345 would (nowadays) be a surviving backup disc for a master nitrate "slow speed disc" (33rpm coarsegroove, usually 44cm diameter), the original of which was sent off many years previously to be made into shellac pressings, cut at 200 Oxford Street London.

In the following sections, I will list the equalisation histories in approximately chronological order. For processed recordings (as opposed to nitrates), I have attempted to give the matrix numbers relevant to each recording system and characteristic. I shall not bother with matrix prefixes, because there were always several in use at once.

6.53 BBC M.S.S. recordings

This section comprises recordings mastered upon the early "Marguerite Sound Studios" (later MSS) equipment between 1935 and 1951. (Ref. 38). The cutting heads were hired to the BBC by inventor Cecil Watts, because he was not happy with their performance and wished to update them immediately. Two frequency test discs of the period survive, enabling us to measure the actual performance today. Number XTR.22 comprises a gliding-tone interspersed with fixed tones. The result is pure "Blumlein 300Hz" from 30Hz to 2kHz, but above that there is a broad peak averaging +3dB from 3kHz to 7kHz, falling to zero at 8kHz, the highest recorded frequency. This disc does not carry any matrix number, so it is difficult to date. But it is a centre-start disc (the tone glides

upwards). Centre-start was abandoned on 27th April 1937, so it is earlier than that, and definitely within Watts' experimental period.

The other is a British Homophone ten-inch test-pressing, almost certainly made to see how British Homophone coped with electroplating nitrate lacquer rather than wax. It has the matrix number BBC210, which would date it to the end of 1935; it carries fixed-frequency tones only. It is pure "Blumlein 300Hz" all the way from 4kHz downwards, although 5kHz is -4dB and 6kHz is -5.5dB. This may partly be due to the low recorded diameter.

Although individual cutters and cutterheads may not have given results quite like these specimens, the basic equalisation characteristic is quite unambiguous – "Blumlein 300Hz."

6.54 BBC American equipment

From 1942 to 1945 the BBC imported large numbers of Presto disc-cutting machines. These would have had a classical "Blumlein shape" until the BBC replaced the original cutterheads in about 1948. Test disc DOM46-2 is thought to date from this period; it carries professional announcements, so was certainly intended for everyday calibration of reproducing equipment, and it therefore shows the *intended* curve precisely. It is absolutely correct "Blumlein 300Hz" up to 10kHz. This assertion is also confirmed by some early coarsegroove 33s, which we shall consider in section 6.56.

6.55 BBC transportable equipment

The BBC's Type A Cutterhead was used on its Type C transportable disc recorders from 1945 to 1961. My graph in section 6.13 shows it recorded Blumlein 300Hz characteristics. Nitrates for immediate transmission cut on these machines generally have an R.P. Ref. No. beginning with M; they are nominally 78rpm, batteries permitting!

The BBC also had portable disc recorders based on clockwork gramophones for its reporters on location during the war. They had piezo-electric cutterheads whose performance has not yet been analysed, although at least one machine survives. However, it seems that most of the discs cut on such equipment were dubbed. (Or worse! Sent by a short-wave radio link, for example). The clockwork motor could only provide enough torque for a ten-inch disc, and that with a relatively shallow groove. As far as I know, very few such recordings were made into pressings for these reasons, although I believe some appeared in the Sound-Effects catalogue.

6.56 BBC coarsegroove 33rpm discs

We have no definite evidence whether the same characteristic was used for long-playing coarsegroove 33rpm records, which became practicable when the BBC started making its own sapphire cutters in 1941. The documentation for the various recording machines does not mention the subject, so one would expect "Blumlein 300Hz" to be valid for these as well. There are a few scraps of evidence to support this claim, but the evidence is not complete.

A rather curious frequency-run appears on the back of some single-sided 16-inch 33rpm BBC Archives pressings (for example X.19910 or X.20865). Presumably it was just a suitable stamper which could be used for the "blank" side. It was definitely recorded on a machine with a manual hand-wound scrolling mechanism (the Presto would qualify). It starts at a diameter of thirteen inches. It too is "Blumlein 300Hz" in shape, but it has a well-damped resonance at 5kHz, and the higher frequencies are much inferior. They peak -4dB at 6kHz and -13dB at 8kHz. I say "peak," because the actual recorded level varies with the rotation of the disc - twice per rotation, in fact, following variations in groove depth. This shows that the mechanical impedance of the lacquer was comparable to the mechanical impedance of the cutter. Neither the MSS nor the BBC Type B cutterheads had such a low mechanical impedance, so I am sure this run was cut by a Presto cutterhead, despite the pressings being made almost a decade later. There are no announcements, so I do not think the run was intended to be used seriously. Instead, I consider it reflects the real performance of the original Presto gear.

Other single-sided sixteen-inch BBC Archives discs have a different frequency run on the back (for example, X18991 or X20755). This is recorded at the full sixteen-inch diameter and includes a professional announcer to speak the frequencies. So it is probably the stamper for a routine engineering test record which hasn't survived elsewhere. But as it has no label, I cannot *assume* this. It is the weak link in my reasoning, which means it might not document the *intended* characteristic. But it is a very accurate "Blumlein 300Hz" at all frequencies up to 10kHz.

The curve for coarsegroove 33s definitely changed at a later date (about 1949), but as 78s are equally affected, I shall present the evidence in the next section.

6.57 Later BBC coarsegroove systems

I am obliged to give this section a rather woolly title, because what I mean is the equipment designed to cut discs to be reproduced by EMI Type 12 pickups. This model of pickup was developed during the last days of the war. Such was the demand from professionals and researchers that it wasn't available domestically for some years. (Refs. 39, 40). With a minor BBC modification, its "open circuit response" (that is, its output when connected to a high electrical impedance) was flat to 10kHz. But if it was terminated with an impedance of the order of ten ohms, a slope approaching 2dBs per octave resulted. This was a reasonable compromise between the constant-velocity of British commercial records and the ideal constant-amplitude curve. I have examined the circuit diagrams of many BBC reproducing systems (including those on cutting lathes) to check that the pickup was always terminated in this manner. It was often used with a matching transformer whose characteristics were described in a different manual, so the circuit diagrams on their own aren't sufficient; but when I took this into account, I found that wherever there was an EMI Type 12 cartridge, there was also a load between ten and twelve ohms, without exception. The 2dBs-per-octave characteristic would be engineered into the cutting amplifiers when the pickups changed.

A new cutterhead, the BBC Type B, was developed which also had a response to 10kHz. It was less sensitive than the Type A, so it could not be used with the Type C battery portable equipment. Its lack of sensitivity (and therefore liability to overload) was compensated by non-motional feedback, which also neutralised a low-frequency turnover due to the resistance of the coil (which now had to be engineered into the cutting amplifiers). The new cutterheads were installed on the new BBC-designed Type D lathes.

Prototype Type Ds were used from the spring of 1945, and aural evidence suggests they were recording Blumlein 300Hz. But there were evidently some modifications before July 1947, when a paper describing the equipment also described the "2dBs per octave" equalisation (Ref. 41), and the production run dated from January 1949. (You can tell a Type D recording because it has a motor-driven scrolling mechanism, giving a runout groove of absolutely constant pitch). The Type D was used for both 78rpm and 33rpm coarsegroove discs, although the radius compensation was different of course. An identical cutterhead mounted in a different shaped case was retro-fitted to the Presto machines about the same time. The electronics of all these machines were designed to give the inverse characteristic to the EMI 12 pickup.

This is documented by two surviving 78rpm test discs, XTR311 and DOM85, both with professional announcements, so giving the *intended* performance[1].

No 33rpm frequency-response test discs for use with EMI 12 pickups survive, but all the circuit diagrams for both recording and reproducing gear show there was no equalisation change between the two speeds, so it seems certain the "2dBs-per-octave curve" applies to 33rpm coarsegroove discs from this time as well.

An exception may be found in Wales, where semi-portable Presto recording kits were fitted with Type B cutters fed from *non*-feedback amplifiers; presumably these behaved like classical electromagnetic cutterheads under these conditions (section 6.13). This would only apply to mobile recordings made by Cardiff, Swansea or Bangor after 1945 (the R.P. Ref. No. prefix would be MCF, MSW or MBG). Also, the wartime Type A cutters continued on Type C lathes elsewhere until at least 1961. To recognise these I use the lack of scrolling facilities, the general instability in groove-depth, the striations caused by the swarf-brush, and (on Current Library nitrates) the M before the R.P. Reference Number. Irrespective of date, these should all be reproduced "Blumlein 300Hz", as we saw in section 6.13 above.

From about 1949 to 1952 the BBC Permanent Library and Sound-Effects sections acquired, and in most cases re-mastered, a collection of wildlife and other sounds recorded by Ludwig Koch. Many of these had first been recorded in EMI's mobile recording van in 1936-1937 using Blumlein equipment, while others were done upon Koch's own portable MSS recorder after the war. Both would have given "Blumlein 300Hz" equalisation. I have had the privilege of hearing test-pressings of some of Koch's originals, and can confirm that they were reproduced to the wrong characteristic when they were dubbed. Fortunately, the situation can be reversed by reproducing the *dubbings* to the "Blumlein 300Hz" characteristic instead of the "correct" characteristic. Ludwig Koch's name always appears on the labels.

Now for the $64000 question. How can you tell whether a recording is done to the "Blumlein 300Hz" characteristic or the "2dBs-per-octave" characteristic? I am afraid you can't. Before May 1945 it is bound to be "Blumlein 300Hz," and after 1949 it is bound to be "2dBs-per-octave," but the changeover is ill-focussed. I can only make the following suggestions:

> (*a*) It would have been logical to change the recording equalisation at the same time as the new pickup cartridges were installed. This was probably done in one BBC building at a time, starting at Broadcasting House and working through the other London premises, through the major regions, to the minor regions.

[1] The curve is rather difficult to synthesise using conventional networks. Adrian Tuddenham has developed a circuit which gives the required characteristic within 0.5 dB.

(*b*) Material recorded especially for BBC Archives, BBC Sound Effects, etc. would have been given higher priority in view of the likely usage at a later date.
(*c*) Listening tests done with a number of BBC Current Library nitrates suggest that the vast majority did not change to "2dBs-per-octave" until the spring of 1949, but the prototype Type Ds may have antedated this.

Some types of disc do not carry any dates - Sound Effects and Transcription pressings - so I will repeat the above paragraph in matrix number terms. Coarsegroove matrixes below 50000 are bound to be "Blumlein 300Hz," and those above 70000 are bound to be "2dBs-per-octave."

6.58 Early BBC microgroove discs

The first BBC microgroove discs of the 1953 Coronation were cut by a commercial firm. A year or two later the BBC commenced its own microgroove mastering on modified Type D equipment, but with new recording characteristics. First we will consider the "domestic" discs (i.e. those *not* mastered by the Transcription Service), which have an R in the matrix prefix. No details of the circuit modifications for the domestic microgroove machine seem to have survived, but there is little doubt that a curve equivalent to "Blumlein 1000Hz" was adopted. (That is, constant-amplitude below 1000Hz - 159 microseconds - and constant-velocity above that). I obtained this information by three different methods: comparisons between BBC Sound Archive discs and surviving master-tapes (e.g. LP25682 and TBS17227), comparisons with commercial discs of the same material (e.g. LP24626 and Decca LF1330), and with double-sided discs of similar material (an example will be mentioned in the next paragraph). I have no doubts myself, but I must stress that this is my subjective judgement. Objective truth cannot be established unless someone discovers a microgroove BBC frequency test disc of the appropriate provenance, or an internal engineering memo on the subject.

This aural evidence also suggests that the "domestic" Type D microgroove machine retained this circuitry long after British and International standards were adopted by everyone else - at least until the end of 1960 - the change being between matrixes 105402 and 105506. Even this is not quite clear-cut, because Take 2s and Take 3s of lower numbers exist, which are also RIAA. One notorious example is BBC Sound Archives LP25926, comprising one long poem read by the author W. S. Graham, which has the matrix numbers 12FRM104536-3 and 12FRM104537. Side 1 is RIAA (the International Standard); but when you change to side 2, there is no readily-available technology which will equalise the sudden loss of bass and treble.

So, if there is *no* take number (implying "Take 1"), if it is a BBC microgroove disc with an R in the matrix prefix, and if the matrix number is less than 105403, I personally consider the item should be equalised Blumlein 1000Hz.

6.59 BBC "CCIR characteristics"

The 33rpm coarsegroove discs made by the BBC Transcription Service in 1954 and 1955 (with a P in the matrix prefix) specifically state "CCIR Characteristics" on the label. The earliest and latest examples known to the writer have the matrix numbers 79342 and 88969. The former was in fact recorded on 23rd December 1953, so it is possible the

Transcription Service intended the new characteristic to take effect from January 1954. The last one cannot be dated, but its subject matter is the Mau Mau Disturbances of February-April 1956. Thus it seems the Transcription Service did not change to either of the new International Curves for their coarsegroove 33s in 1955 or 1956. The turnovers for the CCIR curve were 3180Hz and 400Hz (50 microseconds and 450 microseconds, see section 6.71 below).

The same curve is presumed to have been used for microgroove Transcription discs as well, but it is only occasionally mentioned on the labels. Quite a few domestic BBC Archive microgroove discs were made from Transcription matrixes during these years; these would also be CCIR, unlike the domestic ones. It is easy to tell; the P in the matrix prefix gives it away. Besides, they usually carry a conspicuous sticker over the Transcription label to "convert" them into Permanent Library discs. Or they have words such as "With T.S Announcements" or "As edited for T.S" on the label.

6.60 RIAA and subsequently

From 1956 the BBC Transcription Service used the new international RIAA standard for microgroove discs (75, 318 and 3180 microseconds). This is certainly true for all matrixes with a P in the prefix and numbers greater than 10FPH 97650.

Domestically the BBC did not change to RIAA for its microgroove discs for another four years. The earliest for which RIAA is certain has the matrix number 12FRD105972, but subjective evidence suggests that 10FRM105506 is also RIAA.

There is absolutely no doubt that the BBC continued its own "2dB-per-octave" characteristic (section 6.57) for all coarsegroove discs, whether processed or not. They continued to be compatible with EMI 12 pickups until the last Current Library nitrates were made in 1966.

There are numerous examples of early 78s being dubbed to microgroove in the 1960s and 1970s, and I must warn you that many show signs of having been reproduced to the "2dBs-per-octave" characteristic instead of "Blumlein 300Hz." You may find it necessary to reverse-engineer this mistake, which will only occur on pre-1948 subject matter or location recordings made on Type C gear. If the microgroove dubbing has a matrix prefix incorporating the R which determines that it was mastered domestically, and the matrix number is less than 105403, then the situation has been made worse by the use of the "wrong" LP equalisation as well.

6.61 Brief summary of BBC characteristics

For all pre-1945 coarsegroove discs, and post-1945 ones cut on mobile recording equipment, use "Blumlein 300Hz," except for pressings made from Homo (*sic*) matrixes.

For all post-1949 78s (excluding mobile recordings), and all post-1949 coarsegroove 33rpm nitrates, use the BBC's own "2dBs-per-octave" curve.

For all microgroove records whose matrix prefix includes the R for "domestic," matrix numbers less than 105403, and no take number, use Blumlein 1000Hz. (This applies until about the end of 1960).

For BBC Transcription discs made in 1954 and 1955 (with a P in the matrix prefix), use CCIR Characteristics (50 and 450 microseconds).

For all post-1956 microgroove discs with the P for "Transcription" in the matrix prefix, and post-1961 microgroove discs with the R for "Domestic," use RIAA Characteristics (75, 318, and 3180 microseconds).

For all other discs, please consult the text.

6.62 "Standard" equalisation curves

Instead of studying the industrial archaeology of early disc recording equipment, from now on I shall now be listing the various "standard" curves used in disc recording - in other words those defined in advance, as opposed to those dictated by the machinery.

Such "standards" (apparently clear-cut) are bedevilled with traps for the unwary. I consider the whole subject should be a warning to today's audio industry; practically everything which *could* go wrong *did* go wrong, and it isn't anybody's fault. But much worse is everyone's apparent attempts to *hide* what happened.

6.63 General history of "standards"

The first standards were planned before the advent of microgrooves or slower rotational speeds, and then advertised without clearly saying what they applied to. Immediately after microgroove was introduced, the "battle of the speeds" (33rpm vs. 45rpm) meant that different standards didn't matter, because the discs had to be played on different turntables. But two-speed and three-speed turntables soon became normal, and the chaotic state for commercial recordings began to be realised (Refs. 42, 43 and 44).

At that time standardisation was sometimes opposed on the grounds that microphones, microphone positioning, cutterheads, etc. had a bigger effect (Ref. 45). With present-day knowledge we have an even stronger argument on these lines, although it was never voiced at the time. Record companies made masters on transcription discs or on 78 sides before magnetic tape became good enough and with sufficient playing time; and there weren't any *tape* standards until 1953! Consistent reproduction was only possible from tapes with a full set of calibration tones (with the additional advantage that the performance of the *tape* was documented as well as the recording characteristic).

Other attempts to define disc characteristics suffered the basic ambiguities of definition I mentioned at the end of section 6.7. The Audio Engineering Society of America suggested the "playback half" of the problem, and encouraged engineers to make their recordings sound OK played this way, thus dodging the problem of making a disc-cutting system which actually worked correctly.

Many early LPs were packed in sleeves with equalisation details; but I know cases where sleeves were reissued without the details, or records were re-mastered but packed in old sleeves. I have also met cases where the sleeves were printed in one country and the records pressed in another, with consequent mismatches in the documentation; and second-hand record shops have proved unhelpful, because proprietors (or their customers) swap sleeves and discs to get a complete set in good condition. I can think of only one way to solve these difficulties - collect a number of records of the same make, and of the relevant age and country of origin, and study the sleeves as well as the discs to work out "the originals" (which is what I've tried to do here).

Some manufacturers consistently advocated the same standard curves, even when they pressed records from imported metalwork (or even their own old metalwork) made to other characteristics; and of course there were always "black sheep" who never said anything at all on the subject! There is considerable aural evidence that users of each of the standards converged to reduce the differences between them, although they never admitted it. (Ref. 46)

International disc standards began to emerge in 1955, although leading manufacturers had adopted them somewhat earlier. The microgroove one is commonly called "RIAA" after "The Recording Industry Association of America Inc.", which promoted it. Nearly everyone simply shut up about what they'd been doing to save old stocks of pressings becoming redundant. Such a problem has always afflicted a transition from one characteristic to another of course, but even this begs the question of when a standard "came into force" in the *political* sense. The national standards in European countries all changed at different dates. Any "standard" is preceded by practical tests and experiments, so today we have RIAA LPs which antedated the official promulgation. And of course some time might elapse before records made to new standards actually appeared in the shops. However, for stuff mastered on disc rather than tape, this was better than putting it through two extra generations just to change the equalisation!

We can often say *what* standard a company used prior to RIAA, but it is almost impossible to say *when* the change took place. Straight listening tests aren't reliable when makers were trying to make their records sound like everyone else's.

There were yet more problems after international standards were adopted. Several organisations refused to use them, and at least one country attempted a Unilateral Declaration of Independence on the matter. But the worst trouble was when old recordings were reissued. Although they might be "re-mastered" from original tapes, empirical tweaks might be superimposed to bring them to "modern standards" subjectively. Some were re-mastered with pre-1955 matrix numbers but new "take numbers," and much discographical work, together with spectral measurements and reverse-engineering, may be needed to find the correct "original sound." Or a pre-1955 record would be pressed from original metalwork (or dubbed for reissue) without equalisation being taken into account at any stage, even though the reissue might even carry a post-1955 copyright date. And there seem to be several cases where a manufacturer did *not* follow an unwritten convention – namely the take number on the disc itself should document the attempts to cut a satisfactory disc master (and to process it) until a satisfactory metal negative was achieved.

But the biggest record companies had rigorous engineering procedures, and respected the wishes of their producers. Reissues from-and-by companies like EMI, British Decca, and Philips are usually correct. But for lesser companies, the craft of recognising who cut the master-disc and when (normal to collectors of 78s) needs to be extended to early 45s and LPs. I shall simply be trying to recommend the equalisation to get what the disc cutting engineers intended. This usually does *not* include variables like microphone performance, the performance of a specific cutterhead, or misaligned master tapes. It is assumed that manufacturers compensated these phenomena at the time the disc was mastered if they wished.

6.64 Defining standard curves

I described the RIAA curve in section 6.7. I shall define the others in microseconds as follows. First the high-frequency pre-emphasis turnover; then the turnover between constant-velocity and constant-amplitude (usually equivalent to -3dB points between 300 and 636Hz, or 531 to 200 microseconds); and finally the transition (if any) to constant-velocity again for very low frequencies.

I must also say a few words about where I obtained my information. First, I must explain that I have ignored constructional articles and the like which made presumptions about curves without giving any references to official information, or clues beyond "it sounds best this way." It's not my job to trace such rumours to their source, but as a warning I shall cite Paul W. St. George and Benjamin B. Drisko's article in the US magazine *Audio Engineering* for March 1949. They explained they had made protracted efforts to discover the characteristics used by record companies, but had had only partial success, and were forced to publish a list which they admitted was not necessarily accurate. For example, their list shows (British) Decca's "ffrr" 78-rpm pre-emphasis (section 6.68 below) as 3dB/octave above 3kHz. What is interesting is that this was reproduced in various other articles in American journals until at least November 1954, although I have never found it in Europe, and it is contradicted by "ffrr" test discs and written sources from English Decca themselves. I strongly suspect the source of this evidence was an ordinary tone-control being set to give the nearest flat response and then subsequently measured. And as far as I know, the American company called Decca never used British Decca's curve at all, except when they repressed discs from British Decca's metals (and their catalogues tell you this). And when British Decca established the "London" logo in North America in late 1950, they used the same procedures as in Britain.

Next, other evidence contains inconsistencies and sometimes contradictions, but it usually turns out to be a matter of interpreting ill-expressed evidence correctly. Each characteristic might be documented in one or more ways:

1. In actual time constants. This is unambiguous.
2. Published frequency records, which have to be measured and the results converted to microseconds (often with an element of "curve-fitting.") Since the extremes of the frequency-range might well differ for the reasons I mentioned in section 6.8, I have had to exercise "editorial judgement" to decide where an engineer made a decision to leave the characteristic. I shall use the word "plant" to describe where the master-discs were cut. Usually each disc-mastering plant had only one or two cutting-lathes - sufficiently few for them all to have the same characteristics, anyway. I know only two exceptions: American Columbia, which had at least one microgroove lathe for American broadcasters, and another for the first commercial LPs (which they invented in 1948); and the BBC (who had different procedures for their Transcription Service and their domestic archives - see sections 6.48-6.61above). After this, the problem resolves into determining (*a*) what characteristics were used by what plant at what date, and (*b*) which records (or matrixes) come from that plant. I have attempted to list what I know about British plants, because that is my speciality; but I have only made a very incomplete start upon plants overseas. If anyone has any definite knowledge, PLEASE will they contact me?

3. Published frequency responses, either of the measured response of the actual hardware, or of its theoretical characteristic. Such responses may be in tabular or graphical form. They have to be converted as in (2) above.
4. In the form of circuit diagrams. Theoretically, the stated values of reactors and resistors can be multiplied to obtain the time constants, but (as Ref. 47 points out), source and load impedances can affect the result.
5. In words such as "+12dB at 10kHz." I have *not* used this type of statement, because it is inherently ambiguous.

Next I must state my "tolerances." Most frequency test discs quote a tolerance of plus or minus 0.5dB. Calculated tables are usually given to 0.1dB, but I've found several 0.2dB errors. Graphical representations have been particularly troublesome, because an artist has usually had to hand-draw a logarithmic abscissa and plot a table onto it, sketching intermediate points by hand. Because 6dB/octave slopes don't then look like straight lines, we sometimes have to assume a 1dB error, although the principal abscissae (100, 1000, and 10000Hz) are usually more accurate than that. The values in circuit-diagrams are usually plus or minus 10%, but there are always two such components which have to be considered together. Please bear these tolerances in mind when I describe where I found each piece of information.

6.65 Discographical problems

Having defined the characteristics, the next problem is deciding *which* records of *which* logos were made to *which* characteristics. The correct solution depends upon understanding *who cut the master-disc and when*. This is not the same as the official logo, nor who applied the matrix number, nor where the record was pressed, nor what sleeve it was packed in. For this reason I must reject the noble work of Peter Walker of the Acoustical Manufacturing Co., who provided pushbuttons on his early Quad preamplifiers for correcting different logos. The final version of his work appeared in the Audio Annual for 1968 (Ref. 48). To underline my point, that article gives only one equalisation for "Nixa" (the NAB Curve). But the British Nixa Record Company used American matrixes from Polymusic and Urania (RCA Victor's curve), Westminster (US Columbia's curve), and Lyrichord (three different curves), as well as ones made by itself, all of which changed to RIAA at various dates. I can only recommend that you become familiar with the "look-and-feel" of disc-cutting styles, so you can recognise a master-disc from its "house style," and (more importantly) when an apparent anomaly occurs. In the rare cases where I actually know relevant matrix numbers, I shall give them.

Oddly enough, there isn't much doubt about older 78rpm formats. The difficulty is worst on early microgroove discs, especially those from America, since dozens of logos were being mastered in various plants in competition with each other. However, the average collector using microgroove needn't bother most of the time. It can usually be assumed that if it is a commercial pressing first published outside Germany in 1956 or later, then it will be to the International Coarsegroove Characteristic (50, 450 and 3180 microseconds) if it is coarsegroove, or the "RIAA Characteristic" (75, 318 and 3180 microseconds) if it is microgroove.

Officially RIAA first announced the latter in May 1954, and it was provisionally agreed by the International Electrotechnical Commission in Philadelphia. Many American mastering plants went over to it immediately, I suspect largely because they wanted nothing to do with the broadcasters' NAB curve, or with the ivory-tower AES curve.

However, there were at least seven names for the curve when used in Britain (Ref. 49), where the British Standards Institution published details in May 1955, to take effect from 1st January 1956.

Fortunately, the 1956 British Copyright Act came into effect in June 1957, forcing publishers of records issued in Britain to print the year of first publication on the label, and similar verbiage had to be added to imported records. When this date is 1957 or later and you're handling commercial pressings, you're quite safe; but some nitrates and pressings mastered by small firms had no pre-emphasis *at all* for some years, if ever. However, you may need to research the first publication date for earlier records, which is a bore; and then your work might be rendered useless because the issue might have been re-mastered to the new International Standards at a later stage. There is no unambiguous way I can denote this - you will just have to recognise the "look and feel" of a later matrix by that particular company.

6.66 General history of changeover procedures

I have very little evidence of the type which might satisfy an industrial archaeologist. I have been forced to write this section after a number of listening tests, comparing the same performances mastered in different countries and formats before RIAA, and/or re-mastered in the years after RIAA. Obviously I cannot do this for every case, so I cannot write with great precision. But this "general history" may help readers see there may be several possibilities before 1955, even without the element of subjective "restoration".

Before magnetic tape mastering became normal, the only way of moving sound from one country to another was on "metal parts". In general, the country originating a published recording would send a metal positive to another country. There it could be copied to whatever format the business situation in the destination country dictated. My comparisons show that such metalwork would usually be copied without re-equalisation. Yet I was shocked to discover that the same even applies to American Columbia who had actually *invented* the LP in 1949 (section 6.69). Their format was carefully defined from Day One, because they hoped to set a World Standard. But British Columbia material sent to America may have been joined-up from several metal parts to make a continuous performance; yet it is still Blumlein 250Hz! And I have found a similar situation with US Decca - the latter as late as mid-1954 – and for RCA Victor. None of this even gets a mention in anyone's official specifications.

Shortly after the invention of the LP, American RCA Victor launched the seven inch 45rpm disc, using a much faster auto-changer to minimise the breaks, while allowing customers to choose single items in what later became the "pop" market. The first 45s had a virtually incompatible equalisation standard ("Blumlein 795Hz", or 200 microseconds). It is fascinating to see how other companies tried to match this. Evidently metal-part copying was still in force; some long-playing discs made up from short songs (for example, Broadway shows) seem to be mastered at Blumlein 795Hz. Again, no published information mentions this possibility. Graphs later published by RCA clearly show the 200 microsecond option, but do *not* make it clear it should *only* apply to their 45s!

The next problem is that the commercial recording industry wished to take advantage of the potential advantage of mastering on magnetic tape (cutting and splicing to gain note-perfect performances). But in the absence of "tape standards", no one piece of tape could be guaranteed to play correctly on another machine, unless a fair length of

tape was used for carrying calibration tones to document how well the playback machine matched the recording machine. Two tape standards became available in 1953, one in Europe and one in America (see section 7.8); and because they differed, only the leading professional companies could adapt to the other's tapes. But we can often hear that the quality of early tape was worse than that of metal. So some of the muddles of playing metalwork continued until at least 1954.

Now to the various "non-World standards" for mechanical disc records.

6.67 NAB (later "NARTB") characteristics

The National Association of Broadcasters was an American organisation concerned with the exchange of pre-recorded radio material between various broadcasting stations. It therefore introduced standard equalisation for its members. I do not know when it took effect, but Ref. 50 (December 1941) is the earliest mention I have found, and by January 1943 the recommendations had been adopted (Ref. 51). There were two characteristics at first, one for lateral-cut discs, and one for vertical-cut discs. Obviously they were used by radio stations; but they always seem to have been confined to 33rpm. Initially these were coarsegroove, of course, often on sixteen-inch discs; but some commercial microgroove issues are also known (using the lateral curve). The recording facilities of the Metropolitan Opera House New York also used it, as did some early Lyrichord, Vox, and Haydn Society LPs.

The lateral time constants appear to be 100, 275, and 2250 microseconds. This information is derived from NAB's published graph, which was redrawn for Ref. 51, and a new drawing dated April 1949 was issued by the NAB and reproduced in facsimile in several places. (These graphs are in agreement; but not every writer interprets them in the same way. Please note that the lateral graphs suffer from not having zero-level at 1kHz.)

The hill-and-dale curve is shown in the same places, but shows an extraordinary amount of pre-emphasis, presumably to overcome the even-harmonic distortion of hill-and-dale. The curves actually show the *recording* standard, not the reproducing standard; yet I conjecture the extreme slopes were obtained by inverting a reproducing circuit (probably with an antiresonant top-cut). Nevertheless, very close numerical matching is achieved by assuming conventional time constants at 45 and 550 microseconds, plus two further treble boosts combining to make 12dBs per octave at 9362Hz (17 microseconds) during the recording process. All this was inspired by the development of the first motional-feedback cutterhead (Ref. 52). This gave outstanding performance. It was a decade ahead of its time; but it happened to be a *hill-and-dale* cutter.

6.68 Decca Group (UK) "ffrr" characteristics

The notes in this section apply to discs mastered at the Kingsway Hall or the West Hampstead plant in London. Before the war, "Western Electric" or "Blumlein" shapes apply, as we saw in sections 6.27 and 6.43. During the second world war "Full Frequency Range Recording" was developed. Its full implementation consisted of a new cutterhead, a new microphone, and a new equalisation curve, and they were gradually adopted for 78rpm coarsegroove discs.

The new cutter was available in 1941, after being researched at the behest of the British Government for identifying enemy submarines detected by sonar buoys. Suddenly

a large number of takes called "Take 2" appeared (at a time of acute materials shortages!). Thus I suggest "Take 2" implies the wideband cutterhead, in which case the earliest would probably be matrix DR6570-2 recorded 17th December 1941. But it is thought it ran alongside a non-ffrr cutter for some years, and in any case the full 14kHz bandwidth could not apply until the FR1 full-range microphone had been developed, so it is practically impossible to tell by ear.

However, what concerns us is not the bandwidth, but the point at which pre-emphasis was applied. After the war, Decca's published catalogues started to identify "ffrr" recordings, and it is thought (and aural evidence supports) that these all included the new pre-emphasis. The lowest such catalogue-numbers in each series are F.8440, K.1032, M.569 and X.281, but there are a few non-ffrr items in subsequent numbers, and readers should study the catalogues to identify them. The earliest ffrr session identified this way is Tchaikovsky's Fifth Symphony (the first side of which had the matrix AR8486-2), recorded on 3rd June 1944.

Even so, neither the catalogue numbers or the matrix numbers are foolproof. First we shall consider the catalogue numbers. Decca F.8442 and F.8461 were shown as being non-ffrr, because only *one* side was made to the new characteristic. And I must ask you to be careful if you work with several editions of the catalogues, because sometimes it's the ffrr ones which are marked with a special symbol, and sometimes it's the *non*-ffrr ones.

Next I shall describe matrix numbers for "singles", which are in a common series irrespective of prefixes (AR, DR and DRX). There are several cases of material re-recorded after the changeover, with lower numbers than 8486, but higher take-numbers than "2." All one can safely say is that all numbers between 8486 and 18000 *are* ffrr.

The time constants are 25 and 531 microseconds (-3dB points: 6.3kHz and 300Hz), and this is confirmed by a published frequency disc (Decca K.1802 or London T.4996), and several published graphs.

Other UK Decca logos were affected. UK Decca used NAB characteristics for its own tape recordings, as well as ones imported from America, so that should not be an issue. At first, original American matrixes were imported from US Decca and US Capitol and pressed for UK Decca's logos "Brunswick" and "Capitol", but sometimes these would be dubbed in Britain. (I do not know whether the equalisation was corrected, but of course originals will give better quality than dubbings. This point is currently the subject of listening comparisons). Or copy master tapes would be acquired from America (in which case the equalisation of British-remastered 78s is ffrr and the quality may be higher). When the latter happened with Capitol tapes, they have the matrix prefix DCAP instead of just CAP. Theoretically, British issues of the following Decca-group 78rpm logos are also affected: Beltona, Editions de L'Oiseau-Lyre, London, Telefunken, Vocalion (V1000 series), the very first Felsteds, and taped Decca West Africa series mastered at Hampstead (as opposed to discs mastered on location), until late 1954. Decca also did some mastering for other "independent" British logos: Melodisc, Technidisc, Tempo and Vogue.

When Decca introduced microgroove recording, they introduced a new characteristic also called "ffrr", but which applied to microgroove only. (The coarsegroove characteristic remained as before until International Standards were adopted in January 1956). Historical evidence shows several versions; I shall list them for the sake of historical completeness, although subsequent research has shown this evidence to be very defective, so please ignore the rest of this paragraph! The earliest, in Ref. 53 (Autumn 1950), appears exactly identical to the subsequent RIAA. The second version is dated "Jan. 1951" in Ref. 43; only a few fixed frequencies are stated, but they correspond to

two time constants at about 80 and 680 microseconds. The third version first appears as a "provisional standard" in Ref. 54, which must have gone to press no later than mid-April 1952. Here the author implies there had been too much pre-emphasis in the first version, not surprising when you remember only Decca had a full-range microphone at the time. The characteristic depicted suggests 50, 450, and 2250 microseconds. (But the component values given in his amplifier circuit suggest 66, "may require adjustment", and 2250). A fourth version appears to be documented by frequency test disc Decca LXT2695, which was issued in September 1952. (Matrix number EDP 145-1B on both sides). But the sleeve does not state what the characteristic is *meant* to be called - it may not be "ffrr" at all – although some copies of the disc label show the ffrr trademark. If the levels quoted on the sleeve are supposed to represent the curve, then the closest fit comes from time constants of 48, 300, and 1500 microseconds, while actual measurements of the grooves give a best fit at 41, 275, and 2250 microseconds. In March 1953 a book was published containing yet another version (Ref. 55), but accurate curve-fitting is possible giving time constants of 40, 225 and 1500 microseconds.

My own research shows there were actually three different curves. I did listening tests, comparing ancient LP issues with more-modern versions (assumed to be RIAA), or with original 78s (assumed to be 25 and 531 microseconds). Before I can give the results, I must describe Decca's LP matrix numbers. Prefixes were ARL for 12" and DRL for 10". Next came the matrix number, allocated in numerical order as far as I can tell; and then the take number (signifying the attempts to make a satisfactory metal negative). Then a letter to indicate the disc cutting engineer, and sometimes a W which (I believe) indicates a master-disc cut in wax rather than nitrate. On British versions, an R may follow. This means "Remastered," after which the take numbers go back to 1; unfortunately these may be *any* of the three equalisation curves. However, if the matrix number is *engraved* rather than *punched*, this proves RIAA, because the engraving machine was purchased some months after RIAA was adopted.

The three curves are:

(1) "Blumlein 500Hz" (318 microseconds), used for (Take 1s of) matrix numbers up to ARL1170-2B (and of course many higher take-numbers of these).

(2) 50 microseconds and 318 microseconds, used for (Take 1s of) matrixes ARL1177-1B to ARL2520-2A (and of course many lower numbers with higher take-numbers than 1). It is only fair to say that comparisons sometimes sound better using 40 microseconds instead of 50, but there is no consistency about this. During this period, UK Decca and Telefunken introduced the "MP" format (meaning "Medium Play", ten-inch discs roughly paralleling what had been issued on 45rpm EPs). These matrixes were numbered in the TRL series.

(3) RIAA (75, 318 and 3180 microseconds), used for ARL2539-2A onwards. Decca's RIAA engineering test disc is LXT5346, with matrix number ARL3466-2. MP discs switched to RIAA at about matrix TRL392.

Unfortunately, by the time Decca switched to the new International Standards the company had changed to a new internal procedure for its "single" records. They called it a "Sub-Matrix Number". Each track was given a sub-matrix number when the tape master was made, and this number would be transferred to the 78rpm or 45rpm matrix when the disc was cut, which might be anything up to a year later. Thus we cannot use such a matrix number to tell us when *the master disc* was cut.

The only possible way of breaking this hiatus is to assume that the matrix engraving-machine was introduced simultaneously for both LPs and singles (and EPs and other media), and thus that everything with an engraved matrix number must be RIAA.

Presumably in the case of coarsegroove, this generally implies 50, 450, and 3180 microseconds; but I am told that many late Decca-group *78s* (after about 1958) are RIAA! Evidently there was no change of equalisation – or tip radius - for what were, by then, niche-market issues; but the microgroove versions will have better power-bandwidth product anyway. The 78rpm single with the highest *punched* matrix number seems to be Decca F.10630 (matrix number DRX21213-1), but many singles following this had lower sub-matrix numbers, and others became remastered. The transition to RIAA was certainly before DR21429-2, because that is the matrix number of a 45rpm disc with the catalogue-number 71123, which carries frequencies recorded to the RIAA characteristic. But because such singles lack any chronology, the "grey area" may extend from DR18439 to DR21213.

6.69 American Columbia LPs

First, an important point of nomenclature. I use the pedantic phrase "American Columbia" to distinguish it from British "Columbia", a constituent of Electrical & Musical Industries Ltd., which owned the Columbia tradename in the British Empire and most of the Old World until 1984.

American Columbia was the first organization to make LPs as we currently understand them, and they defined their characteristic very carefully from Day One because they hoped to set a world standard. (They succeeded with the rotational speed and the groove dimensions, but did not quite get the equalisation right!). And they owned the abbreviation "LP" as a trademark, so some writers use this abbreviation to indicate the recording characteristic as well, for example where American Columbia mastered material for other labels.

Since the 1920s, American Columbia had done a great deal of mastering for syndicated broadcasts, so they used NAB characteristics for microgroove as well as their own. (It is possible the deficiencies of the former showed the need for the latter). When they carried out mastering for other logos, they were often NAB. Unfortunately I do not know a rigorous way of distinguishing them, although after mid-1953 (when variable-pitch grooving was used for Columbia's own LPs), the NAB-equalised LPs seem always to be cut at a constant pitch.

American Columbia seems to have allocated consecutive numbers for most of their microgroove masters, starting from 200 in early 1948. By 1954 they had reached about 20000. The prefixes include ZLP and ZSP (seven-inch 33rpm and 45rpm discs), TV (ten-inch 33s), and XTV or XLP (twelve-inch), plus ZTV and XEM ("Epic" seven-inch and twelve-inch) and recordings on the ENTRE logo. However, judging from record reviews and casual listening, I would warn you that material of European origin (Philips and English Columbia) might have the wrong equalisation, resulting in American Columbia LPs with no apparent pre-emphasis. In many cases this can be explained by having been copied from metalwork (section 6.66); but this currently needs further research.

The characteristic was, in engineering terms, a good one, and it does not seem to have changed with time. It was so well-founded that by 1952 constructional articles, texts, and preamplifiers were calling it simply "The LP Curve" without mentioning American Columbia's name at all, which is rather misleading. However, I have yet to see an official statement of the time constants involved. They seem to have been 100, 400, and 1590 microseconds. This information is based on circuits, curve-fitting, and a verbal description of the difference between 900Hz and 10kHz in Ref. 56. But it must also be

said that the graph in Ref. 43 suggests that a slightly closer fit would occur with 100, 350 and 1590 microseconds.

There was a fixed-frequency test disc available as early as September 1949 (American Columbia RD130A), together with a gliding-tone one (American Columbia XERD281) (Ref. 57). The range covered was 50Hz to 10kHz. Another frequency-run to the same characteristic was included on Westminster's LP "High Fidelity Demonstration Record," available in Britain as Nixa WLP5002 (matrix number XTV19129-2AD).

I do not know precisely when the change to RIAA occurred; the differences between American Columbia LP and RIAA are not conspicuous to the ear. Unfortunately, American Columbia itself maintained a public silence about the change. In sections 6.40, 6.67 and 6.71 I gave results of listening tests to resolve such difficulties with other characteristics; but I apologise to American readers for being unable to focus closer with discs I have found on this side of the Atlantic. (Finding two copies of the same performance with different equalisation is practically impossible here). Another difficulty lies in understanding the matrix number suffixes, which do not follow the same system as Decca's "ffrr" (above). In the case of American Columbia, I conjecture that when the subject matter remained the same, the "take number" stayed the same; but if re-mastering proved necessary, the letter following the take number incremented. Anyway, I very much hope someone else may be able to improve my results.

To reduce this difficulty, I researched LPs mastered by American Columbia and issued under the US logos "Vox", "Westminster", and "Haydn Society". These logos did not maintain the aforesaid public silence about the change to RIAA. If the information on their sleeves is correct, and assuming: (*i*) the parent label changed at the same time, (*ii*) the masters were numbered consecutively, and (*iii*) I have not been misled by wrong sleeves, then the changeover seems to be between matrixes XTV19724-1A and XTV19938-1C.

Another theory may be constructed for the material issued under Columbia's own name, whose matrixes have the prefixes LP (ten inch) and XLP (twelve inch). (I do not know if they have the same number-series which follows XTV prefixes, or not). This is that the XLP matrixes switched from (something like) 22000 directly to 30000, so that demonstrators "in the know" would not have to remember a five-digit number in order to apply the correct equalisation.

6.70 "AES" characteristics

This was a *reproducing* curve proposed by the Audio Engineering Society of America in the summer of 1950 (Ref. 58). Curve-fitting gives two time constants only, 63.6 and 398 microseconds, and this agrees with verbal descriptions of the turnover frequencies, and a re-drawn curve in fig. 17.15A of Ref. 43.

As to its use, Ref. 59 (October 1952) also includes a letter from the then-Secretary of the A.E.S, which states that "all Capitol records, and all material recorded by West Coast organisations, is made exactly to this characteristic" (but the statement is certainly wrong for 78s).

He also says the curve was used by Mercury Records; but it does not apply to them all. Early American Mercury LP sleeves carry the message "Changeover from constant velocity to constant amplitude is at 500 cycles and there is a rising response characteristic at the rate of 3 decibels-per-octave beginning at 2000 cycles-per-second." But Mercury had changed to the AES curve by early 1953, and the "Olympian" series (pressed in

Britain by Levy's Sound Studios, often under the logo "Oriole") are certainly stated to be AES, although I have not yet found a pair of examples for a suitable comparison test. To make matters worse, American Mercury metalwork tends to carry a catalogue number (rather than a matrix number), slavishly followed by their UK manufacturing agency, so it seems impossible to state rigorously where the pre-emphasis changed from 3dBs-per-octave to 6dBs-per-octave.

And like other organizations, the AES disguised the changeover to RIAA, by calling it "The New AES Curve" in 1954.

6.71 Various RCA characteristics

Ref. 60 (July 1941) is the earliest contemporary reference I have found which describes RCA Victor using pre-emphasis on its 78s, although the time constant was not given. Straight listening suggests the idea was tried somewhat earlier, and we saw in section 6.23 that Moyer wrote about RCA's Western Electric systems with pre-emphasis at 2500Hz (corresponding to 63.6 microseconds); but I am deeply sceptical. It seems to me far more likely that, if something which had been mastered direct-to-disc was reissued on microgroove, the remastering engineer would simply have treated everything the same. And I consider it likely that judging by "pure sound" clues, Victor's then-unique use of multiple limiters (essentially one on each mike), would itself have resulted in a "brighter" sound.

Between 1943 and 25th February 1955, RCA 33rpm and 78rpm masters can be recognised because the first two characters in the matrix number are the encoded year of mastering. For example, 1943 = D3 and 1952 = E2. (For full details of the prefix system, see Ref. 61). Some were dubbed in the UK for issue by EMI; but apart from that, no other plant has matrix prefixes like this. As with American Columbia, they can turn up in the unlikeliest places.

When RCA Victor adopted microgroove (first with their 45rpm discs, and subsequently their 33s in March 1950), the same pre-emphasis and numbering-system was used, but the bass was cut more brutally. (Ref. 43 shows the same curve being used for all RCA's commercial media in 1954, but maybe this was a hangover from a previous edition of the book). There are only two time constants, 63.6 and 200 microseconds. (No wonder British reviewers complained about the lack of bass on such records). However, listening comparisons between the earliest 45s and 33s show that 200 microseconds only applies to the 45s; the 33s are 318 microseconds. The curve was used for other records mastered at RCA, e.g. Lyrichord and Urania. An "Orthophonic Transcription" curve is also shown, evidently used for sixteen-inch coarsegroove records, with extended treble to the same time constant and an additional time constant equivalent to 1590 microseconds.

It is also known that RCA established a new standard in 1952 which was the same as the subsequent RIAA. (Ref. 62, and see the table in Moyer's paper). RCA called it the "New Orthophonic" characteristic, and these words seem to appear on many LP sleeves to mean that RIAA was intended. Moyer's paper includes a box with the following words: "Use of the "New Orthophonic" curve is recommended for all RCA Victor records and records released by RCA Victor since August 1952. With a few exceptions in the early 6000, 7000 and 9000 series, this applies to all LM, WDM, and DM records above 1701, and LCT and WCT above 1112. It also includes all LHMV, WHMV, LBC, WBC, and Extended Play 45s. Records issued prior to that date should be played *with the same crossover and high-frequency characteristic*, but without the rolloff at low frequencies. A

4- to 5-db increase in response at 50 cps. ... is suggested for these records." With the exception of the words I have italicised in the above sentence, I cannot disagree with it.

A demonstration LP was issued from American Urania 7084, published in Britain in the first quarter of 1954 on Nixa ULP9084. The sleeve clearly shows a RIAA graph, and although there are only five short tones cut on the disc, they are consistent with RIAA. The matrix number is E2 KP 9243, so it dates from 1952. And I am very grateful to Mike Stosich (*pers. comm*) for informing me that his copy of RCA Victor LM1718 is the earliest he has found with "New Orthophonic" on the sleeve; the matrix number on the disc inside is E2 RP 4095.

6.72 "CCIR" characteristics

Roughly, these were the European equivalent of NAB Characteristics, which were drafted in Geneva in June 1951 and agreed in London in 1953 for the international exchange of radio programmes on disc. But they applied to international exchange only (not domestic recordings), and no-one seems to have told those responsible for international exchanges after the domestic recordings had been completed! The same equalisation was used for coarsegroove 78s, coarsegroove 33s, and microgroove records. The time constants were 50 microseconds and 450 microseconds, with no lower turnover.

The 1953 version of the standard added: "Within the USA a different characteristic will be used for the interchange of programmes between broadcasting organisations but the C.C.I.R characteristic will be used by the broadcasting organisations of the USA for international exchange." And the British Standards Institution document of May 1955 (which introduced RIAA for commercial microgroove records) allowed the continued use of the CCIR curve for "transcriptions," defined as "Recordings made for programme interchange between broadcasting organisations and for other specialised purposes and not normally on sale to the public."

A microgroove test disc to the CCIR curve was issued by the British Sound Recording Association at their May 1955 exhibition (catalogue number PR301) (Refs. 62 and 63). But meanwhile the CCIR had agreed to change to the RIAA time constants, and it seems the application of the CCIR curve to "transcriptions" was abandoned very soon after. PR.301 was therefore sold with a conversion table.

I have done some aural comparisons, and found that early (British) Philips and Caedmon LPs used the CCIR curve. But Deutsche Grammophon used 50 microsecond pre-emphasis on its first 33rpm LPs from 1952, with an additional turnover at the bass end of 3180 microseconds. (Ref. 64). There are two ways to identify a 50 microsecond LP: (*a*) if there is a "mastering date" in the label-surround (figures in the form dd.mm.yy), and (*b*) if the label carries the catalogue number in a rectangle. Pressings with 75 microsecond pre-emphasis would carry the catalogue number in an inverted triangle. (This was a German standard).

6.73 "DIN" characteristics

The German Standards Authority DIN tried to introduce another standard, called DIN 45533, in July 1957 (to take effect from 31st October 1957). It had the same equalisation for both coarsegroove and microgroove, although the rest of the world had settled down to having two equalisations. The time constants were 50, 318 and 3180 microseconds,

and at least one stereo test disc was made (Deutsche Grammophon 99105). However, it seems DIN 45533 was not formally adopted, and Sean Davies tells me there was much heart-searching before Germany adopted 75 microseconds instead of 50, to bring them into line with the RIAA microgroove standard.

6.74 Concluding remarks

As early as June 1950, the editor of the respected American journal *Audio Engineering* was bewailing the lack of standardisation. He advocated (and continued to advocate) a flexible equaliser system in which the three time constants could be varied independently and judged by ear. Although all the known pre-determined characteristics are given above, the problem is usually to decide *which* standard has been used on a particular record, and in many cases empirical selection may be the only solution. A suitable system is only available off-the-shelf on the "Mousetrap" disc processor (£3500). The U.S-made "Re-Equalizer" is designed to modify the RIAA curve in a conventional preamplifier via a "tape monitor loop" until it matches one of the above standards; it includes "Blumlein shapes". And, given a constant-velocity amplifier between the pickup cartridge and the main amplifier, it is relatively easy to build a suitable unit. All operators likely to deal with this subject matter will certainly need one of these possible solutions.

Personally, I agree the three time constants should be varied independently. It is then possible to approach the sound you expect gradually, rather than having one switch with perhaps a dozen "standard" curves on it, which must be wrenched around in order to compare the subtle differences between three parts of the frequency spectrum at once. With the three-knob method it is surprising how frequently different operators agree upon one setting, and how frequently the three selected time constants prove to make up one of the "standard" curves.

REFERENCES

1: Peter Copeland, "The First Electrical Recording" (*article*). The Historic Record, No. 14 (January 1990), page 26.

2: Anonymous report of a B.S.R.A lecture by P. E. A. R. Terry (BBC), "What is a Recording Characteristic?", Wireless World Vol. LVIII No. 5 (May 1952), p. 178.

3: "Cathode Ray", "Distortion - What do we mean by it?" Wireless World, April 1955, page 191.

4: H. Courtney Bryson, "The Gramophone Record" (*book*), London: Ernest Benn Ltd., 1935, pp. 86-91.

5: P. G. A. H. Voigt, "Getting the Best from Records" (*article*), London, Wireless World, February 1940, pages 142-3; and in the discussion to Davies: "The Design of a High-Fidelity Disc Recording Equipment" (*lecture*), Journal of the Institution of Electrical Engineers 94 Part III No. 30, July 1947, pages 297-8.

6: Arthur Haddy, "Haddy Addresses the 39th" (*transcript of speech*), Journal of the Audio Engineering Society Vol. 19 No. 1 (January 1971), page 70.

Ref. 7: A. J. Aldridge, "The Use of a Wente Condenser Transmitter" (*article*), Post Office Electrical Engineers' Journal, Vol. 21 (1928), pp. 224-5.

Ref. 8: W. West, "Transmitter in a simple sound field" (*article*), J.I.E.E., 1930 Vol. 68 p. 443.

Ref. 9: Harry F. Olson and Frank Massa, "Applied Acoustics" (*book*), second edition, 1939. USA: P. Blakiston's Son & Co. Inc; London: Constable & Co. Page 102.
Ref. 10: "Notes on Actual Performance Recording": unpublished typescript by F. W. Gaisberg quoted in Jerrold Northrop Moore, "A Voice In Time" (*book*), p. 178.
Ref. 11: R. C. Moyer, "Evolution of a Recording Curve" (*article*), New York: Audio Engineering, Vol. 37 No. 7 (July 1953) pp. 19-22 and 53.
Ref. 12: C. M. R. Balbi, "Loud Speakers" (*book*), London: Sir Isaac Pitman & Sons Ltd., 1926; pages 68 to 73.
Ref. 13: Halsey A. Frederick, "Recent Advances in Wax Recording" (*article*). Originally a paper presented to the Society of Motion Picture Engineers in September 1928, and printed in Bell Laboratories Record for November 1928.
Ref. 14: John G. Frayne and Halley Wolfe, "Elements of Sound Recording" (*book*). New York: John Wiley & Sons Inc., 1949; London, Chapman & Hall Ltd., 1949. Pages 230-1 contain an ill-disguised reference to this disadvantage of the rubber line.
15: H. Courtney Bryson, "The Gramophone Record" (*book*), London: Ernest Benn Ltd., 1935, page 70. (quoting the above)
16: US Patent No. 1527649; British Patent No. 249287.
17: US Patent No. 1637903; British Patent No. 242821.
18: US Patent No. 1669128; British Patent No. 243757.
19: British Patent No. 405037.
20: G. Buchmann and E. Meyer: "Eine neue optische Messmethode für Grammophonplatten," Electrische Nachrichten-Technik, 1930, 7, p. 147. (*paper*). This is the original citation, but the mathematics of the principle were not described until E. Meyer: "Electro-Acoustics" (*book*), London: G. Bell and Sons Ltd., 1939, pages 76-77.
21: H. Courtney Bryson, "The Gramophone Record" (*book*), London: Ernest Benn Ltd., 1935, pp. 85-91.
22: Anon (but based on an I.E.E paper by I. L. Turner and H. A. M. Clark, probably given in June 1939), "Microphones at Alexandra Palace" (*article*), Wireless World Vol. XLIV No. 26 (29th June 1939), pp. 613-4.
23: H. A. M. Clark, "The Electroacoustics of Microphones," Sound Recording and Reproduction (the official journal of the British Sound Recording Association), Vol. 4 No. 10 (August 1955), page 262.
24: George Brock-Nannestad, "The EMI recording machines, in particular in the 1930s and 40s" (*article*), The Historic Record No. 43 (April 1997), note 14 on page 38.
25: Anonymous report of BSRA Meeting presentation by W. S. Barrell of EMI, "Wireless World", Vol. 62 No. 4 (April 1956), page 194.
26: Bernard Wratten, in a letter to Jerrold Northrop Moore dated 24-01-74, quoted in the latter's book "Elgar on Record" (Oxford University Press, 1974) in the third footnote to page 174 of the paperback edition.
27: B. E. G. Mittell, M.I.E.E., "Commercial Disc Recording and Processing," Informal lecture delivered to the Radio Section of the Institution of Electrical Engineers on 9th December 1947. (Available as an EMI Reprint).
28: Microfilms of index cards of EMI metalwork, available for public consultation at the British Library Sound Archive. The HMV metalwork is in two runs on Microfilms 30 to 44 and 138 to 145, and the other logos are on Microfilms 58 to 85 and 146 to 151. Each run is in numerical order irrespective of prefix! Where the cards say "Recorded", this means the date the master disc was cut.

29: F. E. Williams, "The Design of a Balanced-Armature Cutter-Head for Lateral-Cut Disc Recording," (*paper*, received 22nd June and in revised form 24th November 1948).
30: M.S.S Recording Co. Drawing No. CA 13020/A, dated 9th November 1959.
31: Interview with Arthur Haddy, 5th December 1983. The British Library Sound Archive Tape C90/16.
32: Brian Rust and Sandy Forbes, "British Dance Bands on Record 1911-1945" (*book*), page 732.
33: Journal of the Institution of Electrical Engineers, Radio Section, Paper No. 805, pp. 145-158.
34: BBC Engineering No. 92 (October 1972) (*journal*), page 18.
35: Edward Pawley, "BBC Engineering 1922-1972," (*book*) BBC Publications 1972, page 41.
36: ibid., page 42.
37: ibid., pages 178*ff*, 270*ff*, and 384*ff*.
38: J. W. Godfrey and S. W. Amos, "Sound Recording and Reproduction" (*book*), London. First published for internal use by the BBC in 1950, then published for public sale for "Wireless World" (*magazine*) and as a BBC Engineering Training Manual by Messrs. Iliffe and Sons, Ltd., 1952; pp. 69 to 73.
39: "EMI Pickup for Experimenters," Wireless World, Vol. 52 No. 5 (May 1946), p. 145.
40: "High-Quality Pickup," Wireless World, Vol. 55 No.11 (November 1949), p. 465.
41: Davies: "The Design of a High-Fidelity Disc Recording Equipment" (*paper*), Journal of the Institution of Electrical Engineers 94, Part III No. 30, July 1947, page 298.
42: Ruth Jackson, "Letter To The Editor," Wireless World Vol. LVIII No. 8 (August 1952), pp. 309-310.
43: F. Langford-Smith, "Radio Designer's Handbook" (*book*), London: Iliffe & Sons Ltd., 4th edition (1953), Chapter 17 Section 5, pages 727-732 (and supplements in some later reprints).
44: O. J. Russell, "Stylus in Wonderland," Wireless World Vol. 60 No.10 (October 1954), pp. 505-8.
45: Two "Letters To The Editor," Wireless World Vol. LVIII No. 9 (September 1952), p. 355.
46: Edward Tatnall Canby: The Saturday Review Home Book of Recorded Music and Sound Reproduction (*book*), New York: Prentice-Hall, Inc., 1952, page 113.
47: P. J. Guy, "Disc Recording and Reproduction" (*book*), London & New York: Focal Press, 1964, p. 209ff.
48: James Sugden, "Equalisation," in the Audio Annual 1968, Table Two on page 37.
49: John D. Collinson, "Letter To The Editor," Wireless World Vol. 62 No. 4 (April 1956), p. 171. The names he gives are "RCA New Orthophonic", "New AES", "RIAA", "NARTB", "B.S. 1928:1955", "CCIR" and "IEC". Most of these are amendments to existing standards as everyone adopted RIAA, but of course this makes things worse if anything.
50: Wireless World, Vol. XLVII No. 12 (December 1941), page 312.
51: Wireless World, Vol. XLIX No. 2 (February 1943), p. 44.
52: L. Veith and C. F. Weibusch, "Recent Developments in Hill and Dale Recorders," Journal of the Society of Motion Picture Engineers, Vol. 30 p. 96 (January 1938).
53: C. S. Neale, "The Decca Long-Playing Record" (*article*), Bristol: "Disc", Number 15 (Autumn 1950), p. 121.
54: D. T. N. Williamson, "High Quality Amplifier Modifications," Wireless World, Vol. LVIII No. 5 (May 1952), pp. 173-5.

55: G. A. Briggs, "Sound Reproduction" (*book*), 3rd edition; Bradford: Wharfedale Wireless Works, 1953, p. 285.
56: Donald W. Aldous, "American Microgroove Records," Wireless World (April 1949), pp. 146-8.
57: anon., "Test Record List" (*article*), New York: Audio Engineering, Vol. 33 No. 9 (September 1949), p. 41.
58: Editorial, Audio Engineering Vol. 34 No. 7 (July 1950), p. 4.
59: Wireless World, Vol. LVIII No. 10 (October 1952), p. 419.
60: Wireless World, Vol. XLVII No. 7 (July 1941), p. 175.
61: (*Reprints of two RCA internal memos as appendixes to a discography*) The Record Collector, December 1980, page 131.
62: Peter Ford, "Frequency Test Disk and Tape Records", Sound Recording and Reproduction (Journal of the British Sound Recording Association), Vol. 5 No. 3 (November 1956), pages 60 and 71.
63: Wireless World, Vol. 61 No. 7 (July 1955), p. 313.
64: Peter Ford, Reply to Letter To The Editor, Sound Recording and Reproduction (Journal of the British Sound Recording Association), Vol. 5 No. 4 (February 1957), page 101.

7 Analogue tape reproduction

7.1 Preliminary remarks

So far, we have concentrated upon reproducing discs and cylinders recorded by the process of cutting a groove. In this chapter I shall be dealing with the special problems of reproducing analogue magnetic sound recordings. I use the generic term "tape" for all the media, although they include some types of discs, and (in the case of endless-loop tapes) the geometry may be topologically equivalent to a cylinder. Most of my points also apply to the magnetic soundtracks accompanying ciné film and videotaped pictures, and to wire recordings, so my word "tape" will include those as well.

Also, like grooved media, the majority of audio recordings are "baseband", directly accommodating frequencies heard by a human ear (approximately 20 Hertz to 20 kiloHertz). "Hi-fi" video soundtracks are the principal exception, apart from very specialist analogue machines for recording such things as dolphins and bats.

Probably the first point should be "Use a Good Tape Reproducer." This is something which is difficult to define, of course, especially since it seems probable that slow but sure developments will occur in the next few years. Fortunately, it is not difficult to define the performance of a "good tape reproducer" in engineering terms, and your archive should have its machines tested and calibrated until they reach the highest standard. They will then be better than most of the machines which made the original tape recordings, so you will recover the power-bandwidth product without losing anything very significant.

The raw electronics (in "play", but without the tape moving) should have a basic background noise at least ten decibels lower than the noise *with* blank tape moving. The wow and flutter should be less than 0.05 percent weighted on professional formats, and certainly no more than 0.1 percent on domestic formats (this figure is difficult to achieve). The replay frequency-response should match that of a standard calibration tape within plus or minus half a decibel at all the frequencies on the calibration tape, and should do nothing drastically wrong (e.g. have a peak, or a rolloff greater than 6dBs) for an octave either side of those frequencies. The total harmonic distortion on replay should be less than 0.2 percent at all relevant signal levels. (This test is achieved by injecting a magnetic field into the replay head from a figure-of-eight-shaped air-cored coil fed from a power amplifier). It is a lot to ask; but it represents the state of the art for present-day magnetic tape recorders, so it should be easier to exceed this specification in future.

And I must also ask you to be aware that there is always "more in the tape" than we can recover with today's techniques, as we saw in Reference 3 of section 2.8. Thus you should keep the originals in anticipation that more may be recoverable in future.

Today it is normal to post a cassette into a machine or load a reel upon a tape recorder, press the PLAY button, and when recognisable sounds come out we give the matter no further thought. Unfortunately, if we are aiming to reproduce the "original sound" (as mentioned in the Introduction), we must make a great number of checks to ensure we are doing justice to the tape. If we get things wrong, the effect is usually rather subtle - we may not *know* anything is wrong by casual listening - but, as good archivists, it is our *duty* to perform these checks. If we copy the recording to another medium

without making the checks first, we are likely to lumber the people who inherit the copy with a large number of question marks.

From the earliest days of magnetic recording, people were worried by not being able to *see* the sound waves, which had always been possible on disc or film. People were also worried that the magnetism might "fade" with time. Therefore professionals have always made use of line-up tones, recorded at the start of each session to document the performance of the recording machine and of the tape. Such tones may serve to fix the overall sensitivity, the frequency response, and the track identification in the case of stereo tapes. Unfortunately there are no standardised procedures. What I shall be saying in this chapter would be largely unnecessary if every tape carried a set of line-up tones. Unfortunately, bitter experience has shown that we frequently have to work out the levels, the frequency responses, and the track identifications for ourselves. If you come across tapes like this (and everyone does), this chapter will help you nail things down so you will be less likely to make a mistake in restoring the original sound.

I shall therefore relate some points in the history of magnetic recording, and then consider the checks we must perform one-by-one. You do not have to *learn* all this. It is a reference manual, so you can easily refer to it as-and-when you need. But I advise you to read the whole thing through once, so you know the nature of the checks you need to carry out.

If you are an experienced operator, I don't expect you to read it all; instead I shall briefly summarise it by telling you the subject-headings: "Bias", "Magnetised Heads", "Print-Through," "Azimuth", "Standard Equalisation Curves", and "Track Layout Geometry". The special problems of Noise-Reduction Systems are left until Chapter 8, because they apply to other media besides tapes.

7.2 Historical development of magnetic sound recording

There were two principal problems which faced the earliest workers in magnetic sound recording, and until they were solved, the idea remained a laboratory curiosity with low quality.

The first problem was lack of amplification. Even the most heavily-magnetised recording played by the most efficient playback-head gave a very weak electrical signal, barely enough to be heard in an earpiece, and quite insufficient for a loudspeaker. Only when radio receivers using thermionic valves became available did the cost and feasibility of electronic amplifiers result in success for the medium. Nowadays, electronic amplification is so commonplace that we hardly notice it, and the difficulties for the pioneers are almost forgotten. As there are no difficulties in principle today, I shan't consider the amplification problem further.

The other problem has not been completely solved in the analogue domain (and probably never will be). Digital techniques are needed to circumvent it. That problem is the *non-linearity* of all permanent magnetic materials.

By this, I mean the following. As you probably know, a magnetic field can be generated by a current flowing through an electrical conductor. (In practice the conductor is usually wire, and the shape found most effective is when the wire is in a coil, so I shall assume these points from now on). The magnetism generated is directly proportional to the current; but if we try to store the magnetism in a permanent magnetic material, we find that the strength of the stored magnetism is *not* directly proportional to the

"inducing" magnetism. Indeed, the "transfer characteristic" is always a very odd shape. (See Box 7.2)

BOX 7.2
THE "B-H CURVE".

Suppose that the aim is to record the effect of a gradually increasing electric current, for example at the start of a sound wave. As the current begins to flow, permanent magnetic materials resist the magnetic effect (which is natural enough, or they wouldn't be permanent magnets). Only when the inducing current reaches a certain strength does the permanent magnetic material start to respond by taking up some of the magnetism itself.

After that, it absorbs magnetic energy quite efficiently, and indeed its strength rises more rapidly than that due to the coil. This cannot go on for ever of course; as the current in the coil increases, the magnetic material "saturates", and its response slows down again until the material has accepted all the magnetic energy it can. No matter how much more current flows in the coil, the material cannot take any more magnetism. (We may say "The tape is overloaded.")

In practice, sound waves represented as electric currents in wires "alternate"; they go up *and down* in strength. The stored magnetism - hereinafter called "the remanent induction" - suffers further problems when the current in the head starts to fall again. We now have a saturated magnetic tape, and if it is doing its job as a magnetic store, it will resist the effect of the falling current. The electric current not only has to fall back to zero before the material responds significantly, but is actually obliged to flow in the opposite direction before the remanent induction can be pulled back to zero.

Physicists describe the inducing field with the letter "H" and the remanent induction with the letter "B." When the transfer characteristic is plotted on a graph, it is called a "B-H Curve."

When the electrical engineer H. P. Hartley investigated this in the 1920s, he invented a new word to describe the way the permanent magnetism lagged behind what the induction field was doing. He called it "hysteresis", and the term is now used by electronic engineers for any form of cyclic delay, from electromagnets to expense accounts!

The effect upon sound quality is drastic. Even with modern magnetic materials, we get gross distortion of the recorded waveform, so bad as to make human speech practically unintelligible. The problem did not affect the very first Poulsen Magnetophones of 1898, because they were intended for morse signals. (Which proves that digital magnetic recording antedated analogue magnetic recording!) Speech recording only became feasible when Poulsen later discovered the principle of Direct-Current Bias.

7.3 Bias

The solution to hysteresis distortion is "bias." Bias means another signal added to the recording current; so far there have been two forms, "D.C. Bias" and "A.C. Bias."

"Direct-Current Bias" (which is now obsolete) consisted of adding a direct (i.e. steady unidirectional) electric current to the recording-head. Or, what came to the same thing, a weak permanent magnet might be fixed on the recording-machine adjacent to, or inside, the recording coil. This overcame the first problem of the hysteresis curve, its inability to take on the initial rise in current. Indeed, we actually get an amplifying effect, because once this initial problem is overcome, the tape accepts magnetism more readily than the inducing current can supply.

The principle was described thus by Poulsen and Pederson in their 1903 patent: "For rendering the speech more distinct and clear, a continuous polarization current is employed which is led round the writing magnet so as to produce a vibration of the molecular magnets of the speech carrying wire at the moment when they store the speech whereby the clearness thereof is increased." (Ref. 1). It was still necessary to hold the loud signals down in volume to prevent saturation, but DC bias certainly eliminated the worst effects of hysteresis distortion. It was used by nearly all magnetic recording machines until the second World War, and was continued on cheap open-reel and cassette machines until at least the 1970s.

Sometimes two permanent magnets were used: one powerful one to erase any pre-existing sounds, then a weaker one magnetised in the opposite direction to bias the tape to its mid-point. In this way, both erasure and bias were achieved without consuming electric current.

The great disadvantage of DC bias was that *all* the tape was magnetised (whether it was carrying sound or silence), and this meant high background noise. Minute irregularities in the motion of the tape were interpreted by the playback head as sound, and were reproduced as a hissy roar. Furthermore, edits were always accompanied by loud thumps.

"A.C. Bias" was discovered accidentally by some workers in the United States in the 1920s. Their recording amplifier developed an accidental radio-frequency oscillation, and suddenly the recording was almost cured of distortion and background noise. Although their discovery was patented, it was not used commercially. It was rediscovered by German workers during the second World War, and the Allies listening to German radio broadcasts suddenly realised the Nazis had a new high-quality recording system.

In fact, the Germans had combined the "A.C. Bias" invention with the invention of plastic-based ferric-oxide tape. The latter had been demonstrated at the Berlin radio show in 1935, but it was the combination of the two which gave the great step forward.

The principle of A.C. Bias is that an ultrasonic electrical oscillation is added to the signal in the recording head. The frequency of the signal is four or five times that of the highest wanted sound, and cannot be heard. Instead, it "shakes up" the magnetic particles so they can accept the magnetism more linearly, and as the particles are carried away from the head by the tape transport, the effect of the bias gradually dies away. Gaps between sounds result in 'unmagnetised' tape while sounds mean magnetism of the correct strength. Thus low background-noise and low distortion are achieved at the same time.

I shall now describe in words how the addition of ultrasonic bias affects the recorded quality. I do so because you will often come across old tapes which were

incorrectly biassed, and you should be able to recognise the symptoms. There's not a lot you can do about them, but the symptoms can be confused with other faults, which is the reason for mentioning them.

For my description, I shall assume you have a "three-headed machine" (that is, an analogue tape recorder with separate heads for erasing, recording, and replaying), and that you are listening to the playback immediately after the record-head. I shall also assume it is a conventional recording head, carrying both the recording signal and the bias, and not a "cross-field" system in which there are two separate heads either side of the tape, one for the audio and one for the bias. I shall also assume that the machine is set up correctly in all its other parameters, and that you have your finger on the bias control, starting with the bias wound right down.

If this setting of the control results in no bias at all (actually most circuits give a little bias even with the knob right down), you will of course hear the characteristic hysteresis distortion I mentioned in the previous section.

As the bias is increased, the gross distortion diminishes. But the level of ultrasonic bias effectively dies away as the tape moves away from the head, and it has the same effect as a weak erasing signal. This affects the high-frequency content of the recorded sound more significantly than the low-frequency content. If the machine's equalisation has been correctly preset, you will hear too much treble at low bias settings, and too little treble at high bias settings.

The sensitivity of the tape also varies with the bias. It is difficult to judge the effect on actual sound, because the different effects of distorted treble and sensitivity confuse the ear; but with pure tones, the sensitivity (as shown by a meter connected to the playback amplifier) rises to a maximum and falls again as the bias is increased.

The ultrasonic bias also affects the background noise of the tape, and again it is difficult to judge the effect on actual sound. Earlier, I said that the system meant unmagnetised tape during silences; I must now confess this was an oversimplification. In fact, the bias causes the individual magnetic domains to be randomly aligned at the surface, but confined to parallel layers the deeper you go into the tape. The effect varies somewhat with the design of the recording head, but the lowest background-noise is achieved with the bias *not* set for minimum distortion; it comes with rather more bias.

And, if this weren't enough, the point of minimum distortion isn't precisely the same as the points of maximum sensitivity or minimum noise. Thus, it is necessary to do a trade-off between *three* variables in order to bias a tape correctly.

This isn't a chapter on *recording* techniques, it is supposed to be about reproduction; but I cannot leave you in suspense! What modern operators do is as follows. As I said earlier, A.C. bias has a partial erasing effect at high frequencies. But if a particle of dirt should get between the tape and the head, forcing the tape away from the head by a microscopic amount, two of the parameters can be made to cancel each other. The tape receives less bias, so the treble goes up; and the tape receives less audio from the recording-head, so the treble goes down. So a bias setting is chosen in which these two effects cancel, and the results are therefore much more consistent in the real world of dust-particles.

The practice for ferric-oxide tape is as follows. The bias is first adjusted to give maximum sensitivity at a high audio frequency (say 10 kHz), but then it is increased further so that the output due to partial-erasure at this frequency falls by about 2 dB at 38 centimetres per second, or 4 - 5 dB at 19 centimetres per second. (The exact figures are specified by the tape manufacturer, so operators do not have to go through the business of measuring the three parameters on each type of tape). With this setting, the high-

frequency response is usually consistent in the presence of dirt. The equalisation in the recording amplifier is then adjusted to compensate for the fixed high-frequency losses due to the tape-head, the tape, and the partial-erasure effect; and the other parameters are left to look after themselves.

I shall now outline the effects when the bias is grossly misaligned, because this is a frequent problem on old recordings made on amateur machines which haven't been expertly maintained throughout their lives.

Low bias gives excessive treble in the sound on the tape, accompanied by a certain amount of harmonic distortion. This situation usually only arises when there is a severe head-clog, or when a new formulation of tape (requiring more bias) is recorded on a machine intended for an older formulation.

Too much bias means too little treble, but practically no distortion on low-pitched notes. However, the high frequencies are often distorted quite badly, and the subjective effect is known as "S-blasting". The sibilants of speech, for example, are accompanied by low-frequency thumps. This is a very common cause of trouble. As an old-style tape-head wore away, less bias current was needed, because there was less metal remaining; and often such wear was accompanied by an increase in the tape-head gap, so high frequencies also suffered because the magnetism could change significantly as the tape passed across, resulting in another form of distortion. Nowadays these problems are minimised, because tape-heads are fed from a high-impedance amplifier so they draw less current as they wear, while tape-head gaps are better engineered so they do not change their sizes with time. But until the 1960s these problems were very troublesome, and tapes from this period often show overbiassing if the machine which recorded them was no longer brand-new.

The present-day tape operator has no way of curing these faults; S-blasting can sometimes be mitigated by selective low-frequency filtering, but that is all. This means we must both preserve the original and make an "archive copy," because cures for these faults may be developed in future.

7.4 Magnetised tape heads

When the heads of a tape machine with A.C. bias become permanently magnetised, they then behave like heads with D.C. bias, leaving a great deal of background-noise. A similar problem can occur if a reel of tape is placed near a loudspeaker or a microphone. You will also find references in the technical press to losses in high-frequency response and increases in even-harmonic distortion, but I haven't experienced them from this cause.

The result is low-frequency background noise. It is particularly apparent on a tape which has splices, or any creases or unevenness in its surface, because these cause variations in the steady magnetic field, which are picked up by the replay head as thumping noises. Dropouts, or incipient ones which are actually too small to be audible as such, can also cause noises. The effect can also occur long after the original recording has been made, if the tape happens to be run across a deck with magnetised heads or tape-guides. The effect is usually perceived as a series of low-pitched gurgles.

Of course, the solution is never to record or play the tape on a magnetised machine; however, it's too late by the time the archivist hears the effect! It cannot be cured on a dubbed copy. The only way of minimising the noise is to rewind the original tape evenly, perhaps under more-than-normal tension, to flatten out any creases. Experience suggests that splices work better if they're *not* remade. When the tape is

played, it may be possible to fiddle with the back-tension or pressure-pads (if any) to ensure the tape is in intimate contact with the replay head at all times. Another aid is to choose a deck with a reputation for "low modulation noise", a similar fault in which tape hiss goes up and down with the signal. This too is due to irregular tape motion, but is an effect which occurs during the recording process. Although portable "Nagra" machines are absolute pests to rig in a mains environment, you may well find it better to play tapes which have been across a magnetised head on a Nagra (which is beautifully engineered for low modulation-noise).

Otherwise, selective filtering is the only available palliative; this has a subjective element, of course, so cannot be recommended if restoring the original sound is the sole aim. Sections 4.19 and 4.20 described two possible ways of automating the selective filtering on mechanical recordings; but low-frequency noises on magnetic tapes are not strongly correlated with the rotation of the pancake, so building up a consistent "noise profile" hardly ever works. It can be said that (*a*) modern tape machines and tape are less liable to trouble, because of better magnetic materials; and (*b*) some kinds of noise reduction during the recording process (Dolby A, dbx, Telcom, and Dolby SR) can act as insurance to reduce the effect if a tape should become magnetised in the future.

Once again, though, it is the duty of an archivist to preserve the original and/or make an "archive copy" in anticipation of better cures.

7.5 Print-through

This is not strictly a recording phenomenon, but it may cause an archivist a great deal of distress. It happens when the magnetic pattern on one layer of tape "prints" onto the next layer on the reel. The result is "pre-echo" or "post-echo."

I must confess that I haven't experienced much trouble with it myself. Modern tapes have a B-H Curve which is shaped to resist print-through. But you should know that there are several factors which influence the effect, including the following:

1. Print-through increases with time in storage (roughly logarithmically), and with the absolute temperature.
2. Excessively high peak volumes can overcome the initial step in the B-H curve, and cause disproportionate print-through.
3. Physical vibration (e.g. dropping the reel) can increase print-through, as can strong magnetic fields (don't store tapes near a lightning-conductor or a loudspeaker).
4. Print-through is inversely proportional to the thickness of the tape base, and its efficiency is raised when the wavelength of the sound is the same. For practical professional open-reel tapes, human vowel sounds are worst affected.
5. Rewinding the tape before use helps the print-through to self-demagnetise by a few decibels.
6. Storing the tape tail-out when oxide-in (and head-out when oxide-out) takes advantage of the curved shape of the magnetic lines of force to reduce the effect slightly, so that pre-echo is quieter than post-echo (which is more natural).
7. When the tape was recorded, some reciprocal noise reduction processes could have helped, not to mention suitable choice of tape. Unfortunately, tapes with the best inherent power-bandwidth product tend to have worse print-through.

The lessons for storing the tapes are obvious. My concern is "What can we do to minimise print-through on playback?"

For many years, the standard answer has been to power the erase-head of the tape reproducer at a very low level, because the print-through tends to be below the "knee" of the B-H curve and is easier to erase. However, this may also erase some of the wanted signal, especially at high frequencies. Since it can only be applied to the original (not a copy), I consider it important to transfer the tape twice, once before the treatment and once after, to be certain we do not lose anything. I am told the optimum usually occurs when the erasing current cuts the highest wanted frequency (about 15kHz on 38cm/sec tape) by about one decibel.

More recently, Michael Gerzon rediscovered some experiments done by the BBC Research Department in 1953, which showed a better cure was to run the tape past a weak erasing field applied through the *back* of the tape. This does not affect the high frequencies so much, since they are confined to the outer layers on the *oxide* side of the tape. But I have no experience of this process.

A number of workers have suggested various digital signal-processes involving delayed signals and cancellations. It seems to me these would never work properly, because of the hysteresis distortion in the printed signals. But here is an avenue for research. Since it would tackle background noises rather than the wanted signal, any subjective elements would not matter to the archivist.

7.6 Azimuth

Ideally, the gap in a tape head should be exactly at right-angles to the direction of tape travel. If this is not true, the sound at the top edge of the tape will be reproduced a fraction of a second differently from the bottom edge of the tape, and this will affect the reproduced high frequencies. In extreme cases, certain high-pitched notes may completely cancel; this cannot then be cured on a subsequent generation. Another effect is that on a stereo recording, there will be a timing error between the left and right tracks.

The angle of the gap is called the "azimuth angle", which isn't a very good term because that implies movement about a vertical axis, whereas we are (usually) talking about a *horizontal* axis; but that's the name we're stuck with. The azimuth angle should be exactly ninety degrees in all audio applications. (It's sometimes different in video).

If you find a tape with incorrect azimuth, the cure is to deliberately misalign the playback head until its orientation coincides with that of the original recorder. In most cases the error is constant, and can be corrected by ear. It is an objective judgement rather than a subjective judgement, because there will be only one position which gives the maximum high-frequency response. Provided the rest of the machine is correctly aligned to a standard test tape (see section 7.7 below), the correct replay frequency response is assured.

It is perhaps worth noting some incidental points. If the tape was recorded and played back on the same machine, and if the machine had a combined record/replay head (serving both functions), or the replay head was adjusted empirically to match the recording-head, or worse still the recording-head was adjusted to match the mis-aligned replay head (many quality cassette machines work like this), then the engineer would not have heard the effect when he played back the tape. Therefore it is quite normal for tapes to exist with this fault, which no-one knew about until *you* played it!

But it isn't good for your machine to be "groped" unnecessarily. You should not play with the mountings of the replay head just to see whether the azimuth is mismatched or not. Fortunately the solution is simple on most open-reel tape reproducers. As the tape

plays, touch the top edge of the tape with a pencil and push it down. If this is done to the left of the playback head the tape will, in effect, rotate anticlockwise against the head. Repeat the process to the right of the head. If either process results in more treble being audible, *then* it's time to get out your screwdriver and adjust the head-mounting. If the tape is stereo, or you are playing a full-track mono tape on a stereo machine, parallel the two channels (or switch the amplifier to mono) when doing this test. This will double the effect if it exists, and ensures the stereo phase-relationship is optimised as well as the frequency-response. You could also use a Lissajou's display on an oscilloscope or digitisation software, which is even less invasive.

Cassette tapes are particularly vulnerable, and in the author's opinion "azimuth" is the main defect of the format. The difficulties are exaggerated by (*a*) the low tape speed, making the azimuth angle more critical; (*b*) the common use of noise-reduction, which usually exaggerates high-frequency losses; (*c*) the pressure-pad being part of each separate cassette rather than of the machine; and (*d*) the capstan pressure-roller and the pulley downstream in the cassette itself creating significant errors. Unfortunately, it is often difficult to poke the tape as it plays. Indeed, you usually have to remove the loading door to get access to anything. I recommend you keep one dedicated machine for reproducing cassettes if you can, with easy access to a replay head mounted on springs rather than screws, so the head can be rocked slightly by pressing it with a pencil as the tape plays. Then azimuth can be checked; but you should re-check it every time a new cassette is loaded, as individual cassettes may vary, even if they were all recorded on the same machine. (Nakamichi used to make a cassette machine with motor-driven azimuth-setting, and another with two capstans which dispensed with the pressure-pad).

At the British Library Sound Archive, we keep various machines (and tapes) deliberately for experimental purposes. We call them "sacrificial," because (archivally) it doesn't matter if they are wrecked in the process. If you don't use a sacrificial machine, at the end of the session you must always set the head back to the correct azimuth. If it's a combined record/replay head and you don't reset it, all the subsequent tapes made on that machine will be faulty, and you won't know it. I advise you to purchase engineering test tapes and test-cassettes to be quite sure (see section 7.7). They're expensive, but the penalty if you don't have one can be much greater. Reviews in hi-fi magazines show you can never trust even a new machine to be set correctly.

The Cedar "Azimuth Corrector" does *not* correct azimuth! What it does is reduce the time-errors from two tracks on a stereo tape. Therefore you are still obliged to use your pencil and screwdriver to get optimum reproduction of each track.

So far I have assumed the azimuth error is constant, but sometimes it will vary during a recording. It is worth knowing the reasons for this, so you may be able to invent an appropriate cure. One might be that the feed-spool of an open-reel tape recorder has not been placed squarely on the hub, so the pancake is going up and down as the tape leaves it, and the machine has insufficiently precise guides to hold the tape properly across the headblock. The solution is to re-create the fault on a similar machine, although once the tape has been rewound the original relationship between the tape and the spool is lost, and you may have to reset the reel at intervals.

A palliative (not a cure) is to replay only part of the tape track, for example to play a full-track tape with a half-track head or a half-track tape with a quarter-track head (section 7.8). Here the high-frequency response errors are halved at the cost of signal-to-noise ratio. It is a subjective judgement to balance the two disadvantages, and should only be used as a last resort and with suitable documentation to prove you've done your best.

The CEDAR "azimuth corrector" will reduce some of the timing side-effects if the azimuth slowly changes, but it may fail in the presence of loud background noises which differ between the two tracks. (The human ear can detect phase subtleties the machines will ignore). And neither process should be used for stereo tapes without careful thought, *because* they can be considered subjective processes. The CEDAR machine will make its best guess at averaging the time-differences; but spaced-mike stereo involves deliberate time-differences, which the machine may ruin in an attempt to make an "objective copy." This seems particularly important on location events where the recordist has held two omnidirectional microphones at arms' length. All the stereo information depends on which direction he's facing, and if he turns to face something, the evidence that he has changed his orientation is destroyed. Whether this actually matters depends on the application! But in principle I recommend azimuth correctors only for mono, or for true "coincident-mike" stereo recordings.

Also neither process guarantees the full power-bandwidth product when a single track is split into two, because each half of the tape is still being reproduced with some treble loss. I have absolutely no practical experience of what I am about to say, but it seems to me that a logical way to proceed would be to divide each track into four (or eight) sub-tracks, and process them in pairs.

7.7 Frequency responses of tape recordings

This is a surprisingly complicated subject, which isn't helped by the fact that most of the errors which can occur are subtle ones. But it is necessary for me to start by explaining that *all* tape reproducers need some sort of electronic circuitry to bend the frequency response, and the problem is that there are (or were) several philosophies.

The problem arises because most tape playback heads do not reproduce the strength of the magnetism on the tape. (The exception is the "Hall Effect" head, which is not used in audio as far as I know). Instead, audio playback heads have an output which is proportional to the *rate of change* in magnetism they pick up. Since low frequencies change more slowly, they are reproduced more weakly, while "direct current magnetism" (as recorded by a permanently magnetised recording head) results in no output at all until there is a variation, as we saw in section 7.4. Thus low frequencies must be emphasised electronically in the playback amplifier in order to restore the correct frequency balance. The amount of equalisation is large, and that implies a lot of amplification, which (in valve days) was expensive.

There were several philosophies for magnetic wire recording, to judge from the only book I know on the subject (Ref. 2). One was to cover the noise of DC-biassed wire by careful pre-emphasis taking account of the distribution of energy in speech and the subjective perception of distortion; and another was to design an amplifier which could be switched between recording and playback, so it had a compromise equalisation halfway between the ideal recording and replay curves. (There was never a "standard" for wire; the author's experience is that a modern replay amplifier with the same equalisation as 9.5cm/sec tape nearly always sounds right, with only extreme treble requiring attention).

After the initial success of AC-biased tape, anarchic periods followed on both sides of the Atlantic. You will need sheer luck to equalise many of these very early tapes accurately. In the USA, it was thought that the new technology would replace scratchy 78rpm disc records for domestic listening. To make tape players simpler, equipment

manufacturers considered the inherent bass loss should be compensated during recording, and some pre-recorded tapes were marketed on these lines by Webcor.

7.8 "Standard" characteristics on open-reel tapes

However, in American professional circles things were different, because of the dominance of Ampex tape recorders. In 1949, the Radio Manufacturers Association of America proposed "Measuring Procedures for Magnetic Recording" (both 19cm/sec tape and 2ft/sec wire) in the Audio Engineering magazine, which did not mention any equalisation characteristics at all (the recording amplifier being flat within 1dB, thereby requiring thirty or forty decibels of equalisation on playback). (Ref. 3). Constructional articles in the same magazine provided no equalisers for recording-characteristics (only head-losses) as late as June 1952; but by October 1953 the magazine was reviewing a tape-recorder "well suited to the advanced hobbyist who will be satisfied only with the professional type of equalisation." By 1954 Ampex had a standard alignment tape which was available to outside engineers, and in August 1954 the magazine was drawing attention to the two incompatible philosophies on pre-recorded tapes which I mentioned earlier. This was much more severe than the equivalent problem on discs, but it began to be sorted out for the domestic market in 1955.

So, instead of keeping the playback electronics cheap, it seems that from 1955 all American tape recorder manufacturers accepted the principle of "replay equalisation," which is still current today. But to keep things as simple as possible they made it the same at all tape speeds. The standard adopted by the National Association of Broadcasters (the "NAB" standard) was the "50 microsecond" characteristic. Don't worry if you can't understand this; please accept that "50 microseconds" is a quantitative way to define a section of the playback electronics.

However, European manufacturers followed a different path. Broadcasters optimised the circuitry to give the best results for each tape speed. The standards they adopted, known as the CCIR Standards, were issued in the form of recommendations in 1952. The recommendations for 38 and 76 centimetres per second (15 and 30 inches per second) were 35 microseconds, but the case for 19 centimetres per second awaited further discussions. The 38cm/sec recommendation was finally fixed at 35 microseconds in 1953, and the 19cm/sec version at 100 microseconds.

These standards were extended to the magnetic tracks on film and later video. 35 microseconds was used for 35mm film and 15 ips Quadruplex European videotape, and 100 microseconds for 16mm film and 7.5 ips Quadruplex European videotape. The International Standards Organisation proposed these curves for films for the international exchange of television programmes in June 1956. Again, I regret I do not know the dates these standards were actually finalised, but it was certainly done by 1960, and you should use these curves in the absence on any evidence to the contrary.

These standards also became adopted for pre-recorded commercial audio tapes made in Europe. (Ref. 4). A commercial pre-recorded test tape to the 100 microsecond characteristic was marketed by EMI in 1956 (catalogue number TBT.1).

It was logical for someone to propose a 3.75 ips (9.5cm/sec) standard as early as 1955 when such tape speeds became common on domestic tape recorders, and this standard would also have applied to 8mm magnetic film soundtracks; but nothing was decided in "CCIR days." Discussions proceeded in anticipation of a new standard (duly

published by the IEC in 1966), but in the meantime it was anarchy. In practice, curves varying from 100 to 200 microseconds were common.

Now to the crucial point about European and American standards. If you play a professional 15ips CCIR 35 microsecond tape on a NAB 50 microsecond machine, the result is subjectively quite acceptable, but the extreme high frequencies are reproduced up to three decibels stronger. It is not enough to "ring alarm bells" in the casual listener unfamiliar with the original, so clear documentation is called for. But at 7.5 inches per second the difference is much more marked. A 100 microsecond tape played on a NAB machine will sound very "dull", quite different from the original engineer's intentions. The treble is 6dB lower, and the effect spreads down to the mid-frequencies as well. Reciprocal effects occur when NAB tapes are played on CCIR machines.

For completeness, I should also say there is a difference between the two standards at extreme low frequencies. At 38cm/s and 19cm/s, NAB records more strongly below 50Hz (equivalent to 3180 microseconds), so that less equalisation (and therefore amplification) is needed on playback. But I shall ignore this from now on. It might be less significant subjectively, but you should remember side-effects can be considerable if a tape noise reduction system is used.

For British archivists, the problem is made more difficult by the different histories of the record-companies operating in London. The EMI group and the BBC followed the European tradition, while the Decca group (which imported its first machines from the USA) and American-based companies like RCA and CBS used the NAB standard. During the growth in London recording-studios in the 'sixties, it was necessary for British engineers to be able to cope with *both* standards.

In 1967 the 19cm/sec European standard was changed from 100 microseconds to 70 microseconds. This was midway between the two previous standards, so errors were reduced. It also meant that the 19cm/sec changeover frequency was exactly half the 38cm/sec standard, and if tapes were copied at high speed (so that, say, a duplicate was made of a 19cm/sec tape by playing and recording at 38cm/sec), the frequency response along the connecting leads would still be "flat" and measurable on a meter. To take matters further, the 76cm/sec standard was changed to 17.5 microseconds which also gave a flat response along the connecting-leads; the Americans called this "the AES curve." Finally, the 9.5 cm/sec standard was fixed at 90 microseconds and the 4.75 cm/sec at 120 microseconds. This set of standards is known as the IEC Standard. The equivalent equalisations for the magnetic soundtracks on other media running at similar speeds, such as open-reel videotape, 16mm film, 8mm film, and ferric audiocassettes, changed shortly after.

So we now not only have incompatibilities between European and American tapes, but between European ones as well. A pre-1967 tape recorded at 19 cm/sec (or, more pragmatically, such tapes recorded by amateurs on machinery purchased before 1967) will sound "dull" on a present-day machine, affecting the "presence" of the recording. It's only a matter of three decibels difference, but the archivist copying such tapes must make every effort to get the equalisation right.

As the differences are subtle, it can be quite a problem deciding which curve to use in the absence of documentation or frequency tones. Most professional machines could be ordered to any standard, and there have been a few with both European and American selectable by means of a changeover switch (so knowing the make of machine isn't much help). But mass-produced tape recorders for the domestic market were usually made to one specific philosophy. If you happen to know the make of machine, you can eliminate some ambiguities by the lists in Box 7.8, which are compiled from the British *Hi-*

Fi Yearbooks for 1957 to 1981. This shows how makers marketed their machines in Britain; it would not necessarily be correct for the same machines purchased overseas. Also, if you cannot find a name here, it usually means the company did not have a consistent policy from one model to the next.

BOX 7.8.
TAPE RECORDER MANUFACTURERS AND EQUALISATION

MANUFACTURERS SPECIALISING IN NAB MACHINES:
Akai; Baird Varsity; Grundig (not models TK14 and TK18 in 1963); Nivico; Q-Cord; Revox (early transistor machines); Sanyo; Siemens Norge; Sony (except for six models in 1967: TC250, TC260, TC350, TC357, TC530 and TC600); Symphony Pre-Sleep Learning Recorder; Teac; Telefunken before 1970; Uher. (The BBC had hundreds of portable Uher recorders, but they were left as NAB machines - not converted to IEC - as this was considered more satisfactory for speech with the mikes then in use. I would therefore play such tapes with NAB equalisation *unless* the recording is *not* speech, or is known to have been done with an unusual microphone).

MANUFACTURERS SPECIALISING IN CCIR or IEC MACHINES (depending on date):
Abbey; AD Auriema; Alba; Amcron; Bang & Olufsen; Beam-Echo Avantic; Brenell; British Radio Corporation (making machines with the trade-names HMV and Marconiphone); Butoba; Cinecorder; Clarke and Smith; Contronics; Crown; Dansette; Dynatron; Elizabethan; Eltra; Ferguson; Ferranti; Ferrograph; Fidelity; Finex; Fonadek; Gilkes Pakasound; K.G.M; Kolster Brandes; Magnavox; Marconiphone; Philips; Pioneer; Portadyne; Pye; Reps; Revox (valve machines); Robuk; Sharp; Silvertone; Solartron; Sound Riviera; Sonomag Spectone; Stereosound Carnival; Symphony; Telefunken (1970 onwards); Truvox; Ultra; Van der Molen; Volmar; Vortexion; Wyndsor.

MANUFACTURERS WHO MADE MACHINES CAPABLE OF BOTH (switchable):
Bias Electronics; Philips PRO35; Revox (later machines); Scopetronics; Sound Sales; Studer (most models); Telefunken Model 85 de Luxe.

MANUFACTURERS WHO MADE MACHINES CAPABLE OF EITHER (by changing PCBs):
Most professional machines, including Ampex, EMI, and Studer; also Chilton; Ferrograph (model 422); Tandberg (Model TD10); Tape Recorder Developments.

MANUFACTURERS WHO MADE HYBRID MACHINES:
C.W.S Ltd. Defiant T12: "Compromise replay characteristics."
Reflectograph Model S: "CCIR at 7.5ips, NAB at 3.75ips."
Stuzzi: "Close to CCIR."

So far we have been talking about open-reel tapes. Before we move onto cassettes, I should mention that in the mid-1960s there were various attempts to modify the standards to minimise tape noise in professional circles. Ampex, in America, proposed a "Master Recording Equalisation" curve (known as the AME), and Nagra in Europe did something similar with their "Nagra Master" curve. Both were for 38cm/sec only. They may be considered additional to the normal NAB and CCIR curves, so they may be compensated by circuitry external to a reproducing tape machine. They both emphasised frequencies in the range from 1kHz to 8kHz by about 6dB, leaving lower frequencies (which corresponded with louder subject matter) and higher frequencies (where HF distortion set in) unemphasised.

7.9 Standards on tape cassettes

Compact cassettes also suffered from non-standardisation in the very early days, although from about 1970 onwards Philips (the inventors) enforced standards, and this has largely reduced the unknown element. These standards apply to both sides of the Atlantic. Differences are confined to the tape formulation; I shall now explain this.

After the various steel tapes and wires of pre-war years, there were *four* formulations of magnetic coatings on plastic bases. The first was (and still is) acicular gamma ferric oxide. This has been around since the mid-1940s, but has been the subject of several improvements. One was the application of a permanent magnetic field at the time of manufacture to align the magnetic particles before the solvents evaporated (dating from the spring of 1953 in the USA, starting with "Scotch 120"); this gained six or eight decibels of sensitivity, but did not affect the basic signal-to-noise ratio, while the modulation-noise was actually worse to begin with. Next came the use of smaller particles to reduce tape hiss and improve treble (the first manifestation being Scotch 202 in 1965). Smaller particles always mean a more-polished surface, so lower modulation-noise; but the penalty was that fine-grained tape was more susceptible to print-through. (Dolby A noise reduction helped with this). Another modification to ferric tape was to "cobalt-dope" the particles, where the ferric oxide particles take on a coating of cobalt, which has the effect of greatly improving the retentivity. This happened in about 1969. All these tapes are known as "ferric" tapes for short, and when used in audiocassettes they are called "Type I."

Next came "chromium dioxide", first introduced by the DuPont company in 1964 for instrumentation recording. Chromium dioxide is highly toxic, which is why there are only two companies in the world making the basic material today. It has the advantage of shorter magnetic particles, so high frequencies can be accommodated in less tape. So far as the sound operator is concerned, it was first used for video tapes. But in the early days the non-linearity of "chrome" meant there was permanently a few percent of harmonic distortion whatever the signal volume; so it tended to sound "rough", even though the video recorded better. This fault seems to be cured now. "Chrome" is now used in audiocassettes, and they have the classification "Type II"; but chrome has never been used for open-reel audiotape to my knowledge.

"Ferrichrome" cassettes (Type III) were on the market for a couple of years to get the advantages of ferric (low distortion) and chrome (short wavelengths) in one formulation. They comprised a basic ferric formulation with a thin layer of "chrome" on top. They are now obsolete.

"Metal" or "Metal powder" tape, known as "Type IV" when in audiocassettes, was introduced by 3M in 1979, followed immediately by other manufacturers. Its coating comprised particles of pure iron in a suspension. This had the short-wavelength property of chrome, but could hold four or five times as much magnetic strength, with a consequent improvement in signal-to-noise ratio. But to take advantage of this, the recording electronics and heads had to be able to deliver four or five times as much bias and audio without overloading, and even today this isn't always achieved. But metal tapes are ideal for compact video recording (such as the Hi8 format), and are essential for R-DAT digital cassettes. "Evaporated metal" is even more effective, since it comprises pure iron particles or sheet cobalt undiluted by binder; this approximately doubles the magnetic efficiency again, but the manufacturing costs are (as you can imagine) very high. At this point in time, the archival stability of metal-powder tape is doubtful, while evaporated metal tape means a rapid conservation copy must be made, since it often becomes unplayable within a decade.

Now for the point I want to make about playing the different formulations. In audiocassettes, Type I tape is reproduced to a different frequency characteristic from the other three types, and if your machine does not automatically recognise what tape it is, you will have to switch it manually to get the right result. Ferric cassettes (Type I) are meant to be reproduced at 120 microseconds, while "chrome", "ferrichrome" and "metal" formulations (Types II, III and IV) are to be reproduced at 70 microseconds.

This can cause difficulties for pre-recorded cassettes recorded on "chrome" tape (for quality), but intended to be compatible with older machines with only "ferric" capability. So the following system has been adopted.

A second "feeler hole" (next to the "erase tab") is supplied on cassette shells for 70 microsecond tapes, while pre-recorded "chrome" tapes will usually be recorded to the 120 microsecond curve and packed in "ferric" shells. Many newer machines can detect the extra hole and switch the equalisation automatically depending on how the manufacturer intended reproduction to be equalised. (The *recording*-equalisation problem doesn't arise, of course. All pre-recorded tapes should have the erase tab removed, so the consumer cannot accidentally record over the material, compounding the error by using ferric equalisation on a chrome tape).

But if the playback machine has manual switching, the operator must be guided by the presence or absence of the feeler-hole. If you have such a switch, I suggest you play with it while you listen to its effect. It will subtly affect the balance between the bass and treble parts of the frequency range, without actually cutting off any part of the spectrum. It's another example of how playing a tape incorrectly has a *subtle* effect; and for that very reason I urge you to become familiar with it, so you may catch the fault where others might fail.

To complete the story of "feeler holes," a cassette containing metal tape (Type IV) should have a third feeler-hole (at the centre of the cassette-shell edge). This shouldn't affect the playback electronics; it is provided simply to increase the bias and audio power when recording.

7.10 Operational principles

How do you know when a machine is playing a tape correctly? I'm afraid there is no substitute for buying an international standard test tape for each combination of tape width and tape speed. BASF, Webber, and MRL are the principal suppliers.

I would recommend that an archive should make a conscious decision to standardise on either NAB or IEC for its internal open-reel recording work, thereby minimising swaps and mistakes on playback. Whether your archive should keep at least one reproducing machine for playing "the other standard" depends on your workload; but if you do, you will need a second set of calibration tapes. If you are undecided, I would advocate that your collection of machines should be IEC, because it is simple to play NAB tapes through an external passive equaliser patched between the IEC machine and a mixing console It isn't possible to go the other way without (theoretically) an infinite amount of amplification for the bass lift at 3180 microseconds. And for older "CCIR" tapes, you can use a passive equaliser in the connecting-lead to pull up the high frequencies by 3dB.

For audiocassettes I also recommend you get a 120 microsecond test cassette, even if you don't plan to archive on the format.

Machines set up to such test tapes can be used to make "objective copies" without further ado; but if you're lucky and the tape has calibration tones, you should override the alignment of the machine and re-equalise to get the tones correct. (Remember, the tones document the performance of *the tape* as well as the machine).

It is also possible to misalign replay machines to play alien standards by means of the apprpriate conversion graphs. Alternatively, the conversion curve could be set up on a third-octave graphic equaliser, but I have to say it is quite difficult to set the controls accurately enough, and (for theorists) the phase-response is not correctly recovered.

The position for early magnetic film soundtracks is rather different, because there was little interchange between studios until the time came to screen the results, which was not until the introduction of stereo. A 1953 paper on Hollywood stereo film soundtracks showed that at least three equalisation characteristics were in use (Ref. 5). Meanwhile, the EMI stereo 35mm sepmag equipment for the Telekinema at the 1951 Festival of Britain had 40 microsecond equalisation (Ref. 6).

Another difficulty faced by the archivist concerns video tape soundtracks. If your archive is doing its job properly, it may be necessary to separate the video and the audio restoration processes to get the correct audio power-bandwidth product. Certainly, specialist audio machines usually have better wow-and-flutter performances, and the audio can be handled more sympathetically away from the cost and noise of a broadcast videotape machine and the hassles of restoring the video. If the video is being moved to a format with a better video power-bandwidth product, you should do likewise with the audio. Digital video formats which may be used for making preservation copies generally have several audio tracks, and you might commandeer one for reversing signal-compression inadvertently imposed when the video was shot (see chapter 10), thereby providing "archive" and "service" copies on the same tape. This implies that you should know about suitable synchronisation systems as well; but here I advise a manual on video synchronisation methods.

There are so many video tape formats that sometimes archivists have no alternative but to play a videotape on an open-reel audio machine in order to hear the soundtrack (and often to establish what the dickens it is!). This may happen when the tape is in a videocassette, which must therefore be unloaded. The audio will usually be along the top edge of the tape (as you look at the back of the tape moving from left to right). The control-track (which may be essential for setting the speed, or maintaining synchronism in a post-production session) is usually along the bottom edge. With professional video machines (such as Quadruplex and C-Format), the equalisation curve for the linear analogue audio was usually supplied to suit the philosophy of the country

concerned, i.e. IEC in Europe, NAB in America and Japan. But with semi-pro and domestic machines the problem is not so clear-cut. The writer's experience is that such machines used standardised audio printed-circuit boards for the whole world, so for formats which originated in Japan or the USA, the circuitry is NAB. It is quicker to list the exceptions - formats which originated in Europe: Philips N1500, N1700, Video 2000, and Grundig SVC. Of course, if you have the right video alignment tape there's no problem, but this is rare.

If you are lucky enough to have spare time on the correct video tape machine, an audio test tape of the appropriate width, speed, and characteristic could theoretically be run through it to check its audio electronics. But I do not recommend this because it is such a hazardous process; the video machine may *not* run at a submultiple of 38 centimetres per second, and some capstan servos will do strange things if there's no control-track. So the conscientious archivist will make a duplicate of his original audio test tape upon a suitable tape-deck (carefully adjusting the level of each frequency until it matches the level of the original), modify the recording-speed appropriately, and add a suitable control-track along the appropriate edge of the tape, before (possibly) packing it into a suitable cassette housing.

Irrespective of the format, all tapes will have decreased output at high frequencies if the tape has been subjected to appreciable wear. A 1952 paper (Ref. 7) found no appreciable losses after 125 replays, but the level of 15kHz sinewave was -8dB with respect to 1kHz after 1800 replays (probably at 15ips, although this was not stated). Meanwhile, at the Joint Technical Symposium in Paris on 21st January 2000, the Bibliothèque National Française exhibited a statistical survey of videotapes in their possession, showing the remanant induction of their luminance signal (with the highest frequencies) had declined. Tapes made in 1974 had only 41% of their original remanant induction, and this after storage in archival conditions *without even being played for twenty-five years*.

The problem of equalisation is not yet over, because it will be subtly affected if we use a head of the wrong width. (I shall consider track widths in the next section). This is because low-frequency magnetic patterns "leak" across space, and will add to what is picked up by a tape-head which is any narrower than the track. This is called the "fringing effect", and will vary with the tape-speed, and whether the head is against blank tape or air on one side. Normally, test tapes are full width, so you cannot necessarily prove the machine is correctly aligned when you play them with a narrower head.

Quite apart from this, the wavelength of extremely low frequencies can easily exceed the external dimensions of the replay head, which does not intercept all the magnetism, while slightly higher frequencies will cause "bass woodles" (ripples in the replayed frequency-response) at harmonic intervals. This is a playback phenomenon specific to the machine you are using.

Both these effects can be measured by recording low-frequency test-tones onto a machine with the same track-dimensions, and playing this tape on the intended replay machine. (It is comparatively rare for bias and equalisation to affect the *recording* of low frequencies). So far, no-one has developed a circuit for neutralising "fringing effects" and "woodles"; but since they are specific to your reproducing machine, you should carefully explore them and neutralise them with a parametric equaliser if you are making an archive or an objective copy.

Besides the recording and reproducing equalisation of the actual tape, we have to ask ourselves whether we should attempt to undo the characteristics of the microphone in

order to restore the original sound. I have already given my views on this when talking about discs; please see section 6.5.

This seems to be the place to remind readers of a standard "Polarity Calibration Tape" for checking absolute phase (see section 2.10). (Ref. 8). Unfortunately, I know no equivalent for cassettes, video, or films.

7.11 "Mono" and "half-track" tapes

Originally quarter-inch audio tapes carried one channel of sound occupying the whole width of the tape, and therefore the tape ran in one direction, from "head" to "tail". (This is the only system which circumvents the problem of "bass woodles", which I mentioned near the end of the previous section). A similar system is still in use for mono today, when a recording might occupy both tracks of a stereo tape; but this isn't quite the same, because a full-track mono head playing a full-track mono recording will have about 1.6dB better signal-to-noise ratio than a paralleled stereo head playing such a tape. And the signal-to-noise ratio is 1.6dB worse when you use a full-track mono head playing a double-mono recording made on a stereo machine. 1.6 decibels of signal-to-noise doesn't seem very significant, but the tape actually sounds much better than you'd think when the heads and tracks are matched correctly; I suspect modulation-noise may come into it as well.

This shows how careful documentation can improve the power-bandwidth product, even though it may be difficult to hear the difference. Another trick is to use a special head to play the "guard-band" between the two stereo tracks and see if there's anything there, or use a magnetic viewer and see the dimensions of the guard-band.

In about 1948 it was found that, as an economy measure for domestic use, two tracks could be recorded each occupying just under half the width of the tape. This has become known as "half-track" recording (apparently short for "Slightly under half tape-width track recording"!) The recording would start at the "head". This is a word with two meanings when talking about magnetic tape; here it indicates the beginning of the reel. It was usually along the top half of the tape when laced on a conventional "oxide-in" recorder. (Ref. 9). (The exceptions were Grundig 500L and 700L machines of the early 1950s, and the early Truvox Mk. IIIs, which used the *bottom* half of the tape). When the machine reached the "tail," the user would "turn the tape over" (by swapping the feed spool and the take-up spool). Now the machine would record along the other edge of the tape, adding another track. This doubled the time which could be accommodated on a particular reel, and could save having to rewind it.

The trouble was that amateur users proved to be almost incapable of documenting things properly. Tapes may be found with the "head" and "tail" reversed, or the second track might be left blank altogether. This latter is a potential problem for the quality-conscious archivist, because unless he takes the trouble to listen to both tracks, he may not know whether it is full-track or half-track. If a half-track recording is played on a machine with a full-track head, the signal will be about 4.5dB lower while the hiss is the same, so the signal-to-noise ratio will be about 4.5dB worse. Conversely, if a full-track recording is played on a half-track head, the signal will be the same as with a half-track recording, but the hiss will be about 4.5dB higher than it should be, and again the signal-to-noise ratio will be about 4.5dB worse. It's not a *major* degradation, so the operator may not realise he is degrading the signal. So, again, research is essential in the absence of documentation.

When stereo recording began in the early 'fifties, the tape *had* to be recorded with two tracks, *both* starting at the head. If the documentation isn't clear, we must investigate whether the tape is stereo or not, which isn't always as easy as it sounds. Both meters moving together might simply mean it is a full-track mono tape, and record companies sometimes fabricated "fake stereo" versions of original mono master-tapes. Careful listening, or a "sum-and-difference" meter, or an X-Y oscilloscope, may be needed to resolve the difficulty.

In the earliest days of true stereo there was no experience of the effects of crosstalk or timing-errors between the two tracks. We now know that crosstalk figures less than -30dB do not appreciably affect the stereo, but the tracks must be synchronised within a few microseconds. Some early stereo machines reversed these priorities. You are not very likely to come across such tapes, but you should know the principles.

The Bell Labs made a freak machine with two separate tapes fed past two sets of heads and through the same capstan (Ref. 10). A roughly equivalent situation arises in films, where different components of the final soundtrack (such as dialogue, music, and effects) may be recorded on several separate magnetic tracks. These would be pulled through several linked machines by toothed sprockets. The long-playing disc of the first publicly-shown stereo optical film ("*Fantasia*") was made this way, and the microsecond errors are very clear when you switch the LP to mono; but I am not aware this method was ever used for magnetic stereo.

Next, early stereo tape-recording engineers used "staggered" heads - two mono half-track heads spaced a fixed distance apart on the same machine. To reproduce such tapes today, the archivist must use two separate heads *exactly* the right distance apart - within a thousandth of an inch - which requires experiments with one head mounted on a precision movable jig. Alternatively, you could use a conventional stereo tape reproducer and a digital audio editor to alter the synchronisation.

By about the year 1955 stereo tapes were made with "stacked" heads, in which the two channels were recorded by heads placed one above the other in a "stacked headblock." This eliminated timing errors, albeit at the expense of a little cross-talk, and gave mono compatibility in the sense that when the stereo tape was played on a full-track mono machine, acceptable mono sound usually resulted.

7.12 "Twin-track" tapes

But then came *another* complication. In the early 'sixties "twin-track" tape recorders became available, in which two tracks could be recorded *independently*, both starting at the "head". For example, pop music studios could record the instrumental "backing" on one track, and later add the "vocals" on the other, to circumvent problems caused by vocal microphones picking up guitar amplifiers. The complications fork into two at this point.

The earliest such recorders used the *playback* head to feed the musicians' headphones, and the *record* head to make the recording. Thus the vocal and the backing would be in different places along the tape. As these tracks didn't need microsecond-accurate synchronisation, the tape could be moved to another machine (with similarly spaced heads), and the two tracks might be mixed in the disc-mastering room for the published version, thereby cutting generation losses.

More usually, the first track's record-head was operated in reverse as a playback head ('sync head') of reduced efficiency for the musicians' headphones. Generally this

meant inferior high-frequency response and signal-to-noise ratio, with the added complication that sounds from the adjacent stacked recording head would cross the guard-band so the musicians might hear themselves more loudly than they desired. Nevertheless, the resulting two-track tape was in perfect synchronism for a stacked head. The archivist may come across examples of this type of tape, which may sound quite normal when played upon a full-track mono machine, but which actually imply subjective "mixes" of the two tracks.

Tapes like this are often used in the making of television programmes. Because crosstalk is much more critical with unrelated tracks, "two-track" heads are constructed differently from "stereo" heads, with a wider guard-band between the tracks. A two-track tape will play quite happily on a stereo recorder and vice-versa, but a slight increase in hiss (about one decibel) will result. It is usually impossible to tell the difference by ear. If the track width (or the guard-band width) has been documented somewhere, it is possible to choose a suitable replay machine; but otherwise exploration with a magnetic viewer will again be necessary.

As a further complication, television engineers found they could record "timecode" between the two tracks on a two-track recorder (which is the one with the wider guard-band). With suitable interfaces, the tape could then be synchronised with video equipment. When played on a mono or stereo machine with a narrower guard-band (and therefore wider audio heads), this extra track can break through as a continuous chirping or warbling sound.

This is not a monograph about television techniques, but you may also like to know there have been two "standards" for such a timecode - the "Maglink" one (which is analogue) and the SMPTE/EBU one (which is digital). Each can exist in various sub-standards depending on the picture frame-rate; I won't go into this further.

Another twist may be found on American commercial master-tapes of the early 1960s, when background hiss was easily the most conspicuous defect of magnetic tape. The "Dynatrack" system (introduced by 3M) recorded the same sound on two tracks, one to normal NAB characteristics, and the other having an extra treble pre-emphasis shelf amounting to 15dBs with -3dB points at about 900Hz and 5kHz. The latter overloaded horribly at high signal volumes; but upon playback, an electronic switch chose the former. The switch had an "attack time" of 200 microseconds, and when the signal level on the former fell below the 1% distortion level, the circuit switched back in 10 milliseconds. So this is a possible explanation if you come across a tape with a normal and a toppy version of the same material on different tracks.

7.13 "Quarter-track" tapes

To go back to the late 'fifties, the track problem on quarter-inch tape became even more complicated with the introduction of "quarter-track" recording. This is different from "four-track" recording, but the terms are sometimes confused. "Quarter-track" is short for "Slightly under quarter tape-width track", which means that the tape can theoretically accommodate four tracks which may be recorded independently. "Four-track" implies that the recorder can record and replay up to four tracks at once. The "four-track" machine will therefore have four sets of record and playback electronics. A "quarter-track" recorder will have only one set if it is mono, or two if it is stereo.

You can see why amateurs tended to give up with their documentation. To make things worse, there were two different conventions about numbering the tracks on

different sides of the Atlantic. European machines were labelled so that the tracks were numbered in their physical order down the tape. A mono machine would start with Track 1 from the "head". At the "tail", the consumer would turn the tape over, and switch a switch to change to another head three-quarters down the head-stack, which (with the tape reversed) became the second track down the tape. Back at the "head" again he would turn the tape over and leave the switch in the same position, recording on Track 3. At the "tail", he would then reverse the tape once more, and reset the switch to its original position, recording on track 4. For a mono recorder, the switch would therefore the labelled "1 & 4" and "2 & 3". At first, some American manufacturers numbered the tracks differently. What to a European would be "Track 4" would be "Track 2" to an American, and vice versa. If you've followed me so far, you will see that the advantage is that the switch only has to be switched *once* when recording or playing a reel; but the disadvantage is that the tracks are not numbered in their geometrical order. So the track switch would be numbered "1 & 2" and "3 & 4". In April 1965 the NAB introduced a standard which matched the European way of doing things; but this accounts for the difficulty Europeans have in finding their way round American quarter-track tapes, even when they *are* perfectly documented.

The track-numbering for quarter-track stereo is similarly complex, made even worse because it was possible to have mono and stereo recordings on the same piece of tape.

Finally, we have the true "four-track" recorders dating from the mid-'seventies, which tend to be used by home recording studios for four independent lines of music, and keen hi-fi buffs for quadraphonic recording. There is little I can say about these except to alert you to their existence, and to advise you there will be complications in reproducing any of them. Fortunately, the archivist doesn't come across them very often; and when he does, it isn't his job to reduce the four tracks to two or one.

To cope with all these track layouts, I can only recommend that your archive acquires a true four-track quarter-inch machine which will play all four tracks at once. Using this and carefully selecting the tracks on a sound-mixer, it is possible to distinguish between full-track mono, half-track mono in one direction, half-track mono in both directions, half-track stereo, quarter-track mono in either or both directions, quarter-track stereo in either or both directions, and four-track. It still won't unambiguously separate stereo and two-track track-dimensions, but such a machine will greatly reduce the possibilities. It will be especially valuable for finding one item on an ill-documented tape, especially if it can also play tapes backwards without having to turn the tape over. The British Library Sound Archive owns a Studer C274 which offers all these features.

7.14 Practical reproduction issues

Whatever tapes you come across, there might be advantages in having access to a machine which splits a mono track into two (or more). If the tape is damaged or subject to dropouts or head-clogging, relief can sometimes be had from the first stages of the "Packburn Noise Supressor" and "Mousetrap" (see section 0). Besides being able to choose the quieter of two groove walls, they can be switched to select the *louder* of two channels. When the circumstances are favourable, it is therefore possible to mitigate such tape faults by playing a mono track with two half-width stacked heads. Although this may result in short sections of tape having poorer signal-to-noise ratio for the reasons

mentioned above, we are more likely to achieve restoration of the original sound. Equivalent digital processes are easy to conceive.

I must also remind you that you may get errors of *equalisation* if you are forced to play a tape with the wrong head, as well as signal-to-noise ratio. We saw this at the end of section 7.10.

7.15 "Hi-fi" Tracks on domestic video.

Up till now, this chapter has concentrated on applications where the sound is not "modulated" in any way. That is to say, variations in sound-pressure are recorded directly as variations in magnetic strength. Digital pulse-code modulation is another way of recording sound, which I regard as being outside the terms of reference of this manual; but there is a third way, which I shall now consider.

The "hi-fi" tracks of videocassettes are "frequency modulated," like F.M. Radio. Instead of the sounds being recorded directly as variations of magnetic strength, they are recorded as variations in the *frequency* of ultrasonic carriers. Thus, increasing sound-pressures might be represented as increases in the frequency of the carrier-wave. The advantage is that, after demodulation, the sound is largely immune to interference - background-noise - so long as the carrier remains above a certain strength. The process was first used by Sony for their Betamax domestic video-recorders. In 1986 JVC adopted the idea for their rival high-end domestic VHS machines, and this is where the archivist will usually encounter it today.

Conventional F.M theory dictates that the wanted carrier should always be at least twice the strength of other sources of noise in the same frequency-range, whereas "baseband" magnetic recording requires the wanted signal to be hundreds or even thousands of times as much. (This assumes that *speed* errors are not significant; they will add noise otherwise).

All these machines seem to achieve "something for nothing." They modulate stereo sound onto two different FM carriers, and record it "under" the video signal (this is done by recording the audio FM carriers *first*, using an additional head on the video head-drum; then the pictures are superimposed at a higher frequency by a second head whose information does not penetrate so deeply into the tape). In practice, the sound carriers are chosen so they do not impinge upon the carriers for the luminance - which is also frequency-modulated - and the chrominance - which is amplitude modulated - and the tape behaves like the broadcast spectrum with four different broadcasts.

But there is no such thing as a "free lunch," and compromises are involved. The first is that, in practice, you cannot get an adequate signal-to-noise ratio without reciprocal noise-reduction, which is the subject of chapter 8 (but such standards are fixed by the Betamax and VHS specifications). The second is that television sound is generally confined to the frequency-range 40Hz - 14kHz. Substantial amounts of energy outside this range will defeat the noise reduction system, even though the F.M system can easily cope with such frequencies. The means by which television pictures are displayed means the dynamic range of the sound must be compressed (see chapters 12 and 14), and this also helps to drown the side-effects. Finally, you can't edit or change the sound, independently from the pictures.

Yet, compared with a "linear" soundtrack, the subjective improvement is very great. We have the anomalous situation that a cheap VHS videocassette with a "hi-fi" track can have better sound-quality than a professional analogue one-inch open-reel

videotape recorder costing a thousand times as much. This is where the VHS medium comes into its own; on "live" broadcasts, its sound quality might in theory exceed first-generation video tapes in the broadcaster's own vaults.

Now the bad news. Interference always occurs when the video head-drum's alignment is imperfect. If the playback video-head is not running perfectly along the video track, the carrier-to-noise ratio will decrease, and with it the demodulated audio quality. A video dropout may be compensated electronically, but the audio will inevitably suffer a click; and frying noises can appear if the reproduced FM track is not at full level for part of the picture field, or if the head strays off-track. (Fortunately, this is nearly always accompanied by visible picture degradation, so it doesn't happen very often. Professional VHS machines have a tracking meter to indicate the strength of the video signal; the audio is almost always correlated with this. Adjustment of the tracking control should solve the latter difficulty, perhaps in conjunction with other geometrical tweaks familiar to video operators, such as a "skew" control or equivalent).

More troublesome is that domestic formats have two video heads, one for each field of the television waveform. If the signals do not join up seamlessly, the audio carrier is interrupted fifty times per second (sixty times in NTSC countries). The symptom is a "spiky buzz." The noise reduction circuitry is specifically designed to reduce this, although some signals can upset it. Ultrasonic and infrasonic ones are the main problems, particularly on "live shoots" where the signal from the mikes was not filtered. But other signals also occasionally catch it out. An unaccompanied soprano singing quietly in the middle of an opera is the worst case I have met, because the vibrato, the lack of a bass-line, and the relatively low volume made the buzz go up and down very audibly. The same symptoms can occur if either the recording-head or the reproducing-head is slightly worn; as the track is some microns below the surface of the tape, even a small amount of wear or misalignment can capsize the operation.

Cures may require a skilled video engineer to recreate the original "fault", or a brand-new machine without significant wear. Less skilled operators can only tweak the tracking and adjust by ear. It is worth remarking that the *right-hand* stereo channel has the higher carrier-frequency, which is less likely to be replayed cleanly; so (all things being equal) faults may show up more on the right-hand track.

REFERENCES

1: British patent specification 7292/1903.

2: S. J. Begun, "Magnetic Recording," Thermionic Products Ltd., 1949. This is a British reprint of an American book dated June 1948, but I have not been able to find the American publisher.

3: Dr. S. J. Begun, L. C. Holmes and H. E. Roys: "Measuring Procedures for Magnetic Recording" (proposed standard), RMA Subcomittee on Magnetic Recorders R7.4, published in Audio Engineering Vol. 33 No. 4 (April 1949) pages 19 and 41-45.

4: "HMV Tape Records," Wireless World Vol. 60 No. 10 (October 1954), p. 512.

5: Discussion after delivery of paper by J. G. Frayne and E. W. Templin, "Stereophonic Recording and Reproducing Equipment," Journal of the SMPTE, Vol. 61 p. 405 (September 1953).

6: Wireless World Vol. LVII No. 6 (June 1951), page 223.

7: W. S. Latham, "Limitations of Magnetic Tape," New York: Audio Engineering Vol. 36 No. 9 (September 1952), pages 19-20 and 68-69.

8: Stanley P. Lipshitz and John Vanderkooy, "Polarity Calibration Tape (Issue 2)", Journal of the Audio Engineering Society, Vol. 29 No. 7/8 (July/Aug 1981), pages 528 and 530.
9: British Standards B.S. 1568:1953 and B.S. 2478:1954. "Magnetic Tape for Domestic and Commercial Recording."
10: S. J. Begun, "Magnetic Recording" Thermionic Products Ltd., 1949, page 156. This is a British reprint of an American book dated June 1948, but I have not been able to find the American publisher.

8 Optical film reproduction

8.1 Introduction

The idea of recording sound on optical film is so closely associated with the idea of recording moving pictures photographically that it is practically impossible to separate them. I do not propose to try. All I can do is to list some techniques which sound operators in other genres have up their sleeves, which could be applied to optical tracks (and *should* be, if we agree to "restore the original intended sound").

Before I do that, I shall mention two or three optical recording formats which were intended for sound-only applications. These did not survive long because of the costs. Photographic systems, such as the Selenophone and the Tefi-Film, required silver-based chemicals, which were coated onto bases requiring precision engineering techniques, and suffered from the further disadvantage that the soundtrack could not be replayed (or even seen) until the film had been developed much later.

The Philips-Miller mechanically-cut optical film did not have the processing lag, but it was confined to areas where mass production could make no contribution to reducing costs, and it so happened that the films were on a cellulose nitrate base. The originals now seem to have been scrapped, either because they were degrading, or because of the fire hazard. I do not know of any original Philips-Miller recordings surviving anywhere; we only have tape and LP transfers. It is possible, of course, that some of the techniques for optical sound restoration might still be applicable to these. Details of the mechanical cutter may be found in Ref. 1.

But even when we *have* a sound-only optical format, it is the author's experience that it is better to make use of the expertise and equipment of a film-based archive to be certain of reproducing the power-bandwidth product in full. As I said in my Introduction that I shall not write about esoteric media, I shall just describe the soundtracks accompanying moving pictures.

8.2 Optical sound with moving pictures

Oddly enough, the history of moving films has always meant that sound has taken second place. I do not mean to say that sound *recording* was a neglected art - although it sometimes was of course! - but that sound *reproduction* was generally made "idiot proof". It makes the job of the film archivist easier for two reasons. Firstly, film sound engineers usually took reproduction difficulties into account when they made their soundtracks. Secondly, standards were maintained with few alterations, otherwise the "idiot-proof" advantages would be lost.

On the other hand, if we are working in an environment where we wish to restore the original sound (rather than the original *intended* sound), then it is a different matter, because the working practices of contemporary engineers were largely undocumented. This problem might occur if we are taking bits of optical film soundtrack for sound-only records, or if we are working for the original film company and they have changed their practices.

Thousands of improvements were proposed to the idiot proof methods over the years, but they were all essentially experimental in nature, and very few broke the mould. As I said I would only concentrate on mainstream media, I feel absolved from listing the experimental setups. As far as I know, all these experimental films can be tackled by an experienced operator with a thorough grounding in the basic principles listed here and elsewhere, in the same way as an operator experienced with grooved media can recover the full power-bandwidth product from alien cylinders and discs.

From now on, I shall be making frequent reference to different "gauges" of optical films. Gauge is always quoted in millimetres, and includes *both* the picture and the sound when they are combined onto the same print.

8.3 Considerations of strategy

Archivists are hampered by the fact that there is no satisfactory non photographic method of storing the pictures. As I write this (early 2001), we are a very long way from being able to convert the best moving picture images into electronic form, much less digital form. Indeed, we can only just envisage methods of dealing with "average" quality film pictures - say standard 35mm prints - and even with this material, the apparatus isn't actually available yet. I am talking now about "archive" and "objective" copies. It remains perfectly feasible to use quite inexpensive video formats for "service copies."

As a sound operator, I consider the prime difficulty is simply that no electronic method can give pictures to the speed accuracy normal in audio. Until this problem is solved, I cannot see that it is worth bothering with the pictures! But, as someone who has had to work professionally with film pictures, I can see that questions of definition, contrast range, field of view, etc. also desperately need attention.

Even if these problems were to be solved, we would probably not be able to reproduce the images in the manner intended by the original makers. An adequate "screen" for electronic display of film pictures is only just beginning to appear (the flat screen plasma or LCD display); but this will remain peripheral for the following reasons.

Films were mosty exhibited in cinemas and "peep-shows" before the arrival of sound, and a standard running speed wasn't needed because the human eye is generally more tolerant of speed errors than the human ear. As projected pictures grew brighter, faster frame rates became desirable because the human eye's persistence of vision," which is shorter in brighter light. Average frame rates gradually climbed from 12-16 frames per second in the early 1900s to between twenty and twenty-four by the 1920s. When the "talkies" started, it was a natural opportunity to make a clean break with former practice, and upgrade and standardise at the same time. The result was 24 frames per second. Later developments have shown the choice to have been very wise - it is practically the only speed at which you can shoot under both 50Hz and 60Hz mains lighting). So when 24 frames per second was adopted with sound, that was it - further changes were not possible without upsetting the sound.

The boot was now on the other foot. National standards authorities recommended a peak screen brightness for indoor cinemas in the neighbourhood of 10 foot-lamberts. (Ref. 2). This was done to minimise the motion-artefacts for the cinemagoer. It was also assumed that the screen would occupy most of the field of view of the observer with no surrounding visual distractions, and there would be no stray light to fog the image and reduce the contrast range, which might otherwise be as much as 50: 1.

However, electronic displays were developed for domestic viewing (television). This meant pictures many times brighter than 10 foot-lamberts. Under these conditions the persistence of vision is shorter, and motion becomes appreciably jerky (especially on a big screen). To ameliorate this, television's brightness range rarely exceeds 10: 1, otherwise details may be lost because of stray light from the surroundings. Domestic viewers will never put up with having their rooms stripped of visual distractions, nor will they put up with a screen which occupies most of one wall. So material shot for television has a faster pace of editing, more camera movements to provide instant recognition of depth using parallax, larger screen-credits, and frequent closeups of characters' faces. With no consumer base, the finance for developing suitable processes for cinema-style images in the home seems unlikely.

For this reason, most film archives have kept photographic film as the destination medium for their restoration work. It is a well known medium. It can be projected easily to paying audiences in an idiot-proof way, and converted ("downgraded") to whatever electronic medium users might demand. Its costs are high but well known, and it is a much more stable platform to work from than rapidly shifting video technology. It is not a format with great longevity; but the problems are well understood and the timescale is predictable. Unhappily, the technique of copying photographic films onto photographic films gives difficulties with the soundtracks, because they are essentially being copied from an analogue medium to the same analogue medium, hence potentially halving the power-bandwidth product.

Controversy even rages on this point. Modern optical film soundtracks are capable of slightly more power-bandwidth than many early ones (especially those made before 1934). Should these be raised in volume to fill the power-bandwidth product of the destination medium to reduce the losses, and make them ready for screening without rehearsal? Or should the original tracks be photographically copied as they are (warts and all) to preserve the style of the soundtrack unadulterated? Unfortunately, the cost of picture copying and storage is so high that it is usually impossible to afford *two* versions, "service" and "objective" copies to use our language. This, in turn, means that if we get the ideal high-definition variable-speed digital video format and its screen, it will also have to hold many alternative synchronous PCM digital soundtracks.

Because of all this, I make no apology for the ivory tower nature of the following sections. I shall simply make recommendations for recovering the power-bandwidth product of optical film soundtracks, without any reference to how the results might be stored or displayed.

8.4 Basic types of optical soundtracks

There are many types of optical soundtrack[2]. The simplest is the variable density type and its equivalent is the variable area or variable width type. When either of these passes through a narrow uniform slit of light shone across the track, its brightness is modulated by the film, and the remaining light can be converted directly into an electronic audio output. The device which performs this task has had many names. Until the 1960s, it was always a "photo-electric cell" or "photocell"; but nowadays we use the catch-all term 'photoelectric devices'.

[2] Figures showing different optical soundtrack configurations including a modern stereo variable-width track intended to illustrate this section are not available.

Somewhat more complex types of soundtrack are a dual variable-width track, in which the net effect is the same as the variable area type but the modulation is split into two halves. In the pushpull track the two halves are out of phase. This cannot be reproduced with a single photoelectric device. The scanning beam has to be split into two and fed to two separate photoelectric devices, and the output of one of them reversed in phase electrically. Many sources of even harmonic distortion are cancelled this way.

In some types of variable width track, in the absence of sound, the clear area closes down to a thin narrow line. Thus the effects of dirt and halation are reduced; but it is necessary for the timing of the circuitry to be carefully optimised to eliminate thumping noises and the breathing effects following transient sounds. Again, a variable density equivalent exists, which also has the advantage that a limited amount of volume changing can be done during the film printing process if required. With a push pull equivalent of this format the thumps are automatically cancelled, and the noise reduction can therefore be faster.

In a Class B push-pull track the positive going and the negative going halves of the wanted sound are separated. This gives optimum immunity to dirt whilst using a single photoelectric device, but has the risk of crossover distortion at low signal volumes, and increased distortion at all signal levels if the illuminating beam is not uniform and on-centre.

The advantages and disadvantages of most of these formats were worked out during the first twenty years of sound on film, and are described at great length in Reference 3 and elsewhere. The sound restoration operator is obviously helpless to do anything about these types as they come to him, except understand the disadvantages and how they may be minimised (e.g. by ensuring the beam is uniform, the print clean, the dual photoelectric devices are in place, etc).

8.5 Soundtracks combined with optical picture media

Because cinema pictures undergo intermittent motion in the projection gate whilst the soundtrack must run at constant speed, the sound is displaced with respect to the picture. On all optical sound formats it is nearer the head of the roll. As the film usually travels in a downwards direction when it is being screened optically, the sound head is placed *beneath* the picture gate. The intervening film forms a flexible loop while a flywheel stabilises the speed at the sound head. Compliant and resistive mechanical parts (springs and dampers), and power assisted mechanisms, may be provided to allow the flywheel to get up to speed in a reasonable time without the film being scratched. All this needs critical levels of care and maintenance to avoid wow and flutter, particularly noticeable on narrow-gauge formats.

Students of the history of technical standards for 35mm film will find that the amount of this displacement has apparently varied, with 19 frames (14.25 inches) and 20 frames (15 inches) being quoted. It wasn't until the ISO Recommendation of December 1958 that this was made clear. There, it was set at 21 frames, and an additional paragraph explained that when the distance was 20 frames, the picture and sound were in synchronism for an observer at a distance of 15 meters (50 feet) from the loudspeaker.

In theatre projection circumstances, these figures might be altered to allow for the time taken for the sound to travel from the loudspeakers behind the screen to the middle of the auditorium, or perhaps to the most expensive seats (which tended to be even further away). The size of the lower loop was the usual way of regulating this; so if you

are taking the sound from a conventional projector and moving it to another medium, this displacement should be checked and, where necessary, allowed for.

On 16mm films the nominal difference is 26 frames, and on super-8mm films it is 22 frames.

It is worth noting that all three gauges may be found with magnetic stripe instead of, or in some cases in addition to, the optical soundtrack. In an attempt to keep to the standards, early workers with magnetic stripe kept their displacements the same as optical film. Towards the middle of the 1950s magnetic sound-on-film became dominant for handheld shooting, such as news film. Because television news programmes frequently needed to mix optical and magnetic stories on the same reel and transmit them without relacing, the magnetic heads were thereafter displaced by different amounts: 28 frames in the case of 16mm and 18 frames on super-8mm.

Magnetic stripes for 35mm and larger gauge formats were generally used with wide-screen spectaculars having stereophonic sound. Because the conventional projector layout could not accommodate extra heads in the usual place, they were mounted *above* the picture gate. On fourtrack 35mm Cinemascope prints, for example, the distance is minus 28 frames.

All the standards authorities agreed that the tolerance should be plus or minus half a frame, and that the measurement should be from the optical slit (or magnetic head gap) to the middle of the appropriate picture frame.

This writer has many times been faced with the question, "What if it *looks* wrong?" A skilled film editor can easily judge less than half a frame on critical scenes. (This isn't difficult to comprehend when you realise that orchestral musicians can consistently follow a conductor with an accuracy approaching a hundredth of a second). This may mean separate archive and service copies.

8.6 Recovering the power-bandwidth product

As with any analogue recording, the basic principle is that the nearer you are to "an original," the better the quality is likely to be. There are, however, a couple of special caveats for optical film.

The first is that the original may often be an optical negative. If this is the case, the principle of the narrowed track - preventing the reproduction of dirt - is stood on its head, and it no longer works. However, the high frequency response and the harmonic distortions will be better on the negative, and printing to positive stock will lose some of this information. I have no practical experience of my next suggestion, but as a sound operator I consider it would be worthwhile reproducing *both* the negative and the positive print, and combine them using the digital equivalent of the process outlined in Section 0.

I must also record that Chace Productions Inc., of Burbank, California, have a process for scanning an optical negative digitally. This allows the process of ground-noise reduction to take place in the digital domain; but I have no experience of the results.

Where there is no narrowed track (or its variable density equivalent), the maximum power-bandwidth product is bound to exist on the negative. (Or, to be pedantic, the original positive in the extremely unlikely case of a comopt reversal original). But even-harmonic distortion will occur because the "gamma" (the linearity of the relationship between the brightness and the desired output voltage) will normally have been controlled in the printing and development stages. However, provided the negative is

transferred as it is without any phase shift, even harmonic distortion can (in principle) be corrected in the digital domain.

An optical soundtrack may be copied to another in three ways. Contactprinting holds the master and the unexposed films in close contact and shines light through one onto the other. This gives perfect transient response, of course; but it is not often done, because the films have to be held in close registration by their sprockets, and this can increase flutter. Optical printing shines light from one film running on one continuousmotion transport through an optical system directly to the unexposed film running on another continuousmotion transport. Various aperture effects can cause high-frequency losses, and careful sensitometric testing is essential to avoid harmonic and intermodulation distortions, but the system is sometimes used for copying sound from one film gauge to another. When used for this purpose, the frequency characteristics are not necessarily optimised for the destination medium, so the third method is preferred (although much more expensive). This is called electrical printing, and simply means that a reproducer is connected to a recorder using an electrical connection (the same means as copying magnetic tape). Thus aperture effects may be compensated at the expense of putting the sound through twice as many optoelectronic and electrooptical transducers. Frequency re-equalisation and possibly dynamic compression may be employed to give better results, especially if the destination medium is a narrower gauge; but the full disadvantages of analogue copying (doubling the distortion, noise, and speederrors, and introducing subjectivism) can occur.

The soundtrack is usually scanned by a narrow rectangular beam of light, corresponding to the gap in a magnetic tape head. Because the film is usually moving downwards rather than sideways, the dimensions are called height (the narrow one) and width (the wide one). It is the height of the beam which plays the most important part in the playback frequency response. Accurate focussing and azimuth also play a part, but for many years there was a constant battle to get enough light to operate the photoelectric cell without backgroundnoise coming from that component. All other things being equal, the better the highfrequency response, the worse the photoelectric noise. Nowadays, solid-state photoelectric devices have noiselevels near the thermal limit and there is less of a problem. We can therefore concentrate on extending the frequencyresponse, and the penalty of noise will be due to the film itself rather than the playback mechanism. Which is the otcome expected by archivists.

The light should be focussed on the emulsion side of the film, and if the optical system is welldesigned, scratches and dirt on the clear side of the film will be defocussed and will give less output.

The slit height is typically 0.5 to 1 mm. At this sort of size, significant amounts of optical diffraction can take place, especially towards the red end of the spectrum; and blue light (or even ultraviolet light) is to be preferred if there are no disadvantages.

Azimuth is much more critical than with analogue tape. Fortunately, severe azimuth errors at the recording stage are uncommon, and conventional idiotproof projectors do not even have an azimuth adjustment. It isn't just high frequencies which can be affected, although this might be very significant with narrower gauges. Very considerable amounts of intermodulation distortion are also generated. This usually comes as a surprise to operators used to tweaking azimuths on analogue magnetic tape; but because films are recorded to constantamplitude characteristics (as we shall see in the next section), extremely steep waveforms may be encountered. Some highpitched waveforms on variablewidth tracks may slope at an angle of 89 degrees. A mathematical description of how the distortion will occur on three types of soundtrack may be found in

Ref. 4; but I have found it helpful to describe the problem to audio engineers by asking them to consider what would happen if you played a lateralcut disc with 89 degrees of tracking error!

This seems to be the place to remind operators that the equipment must have low levels of distortion itself, besides the extended frequency-response and low noise. Subjective experience backs this, particularly when the soundtrack is poor. It seems that under these conditions the ear needs all the clues it can get to perceive the recorded content, and relatively small amounts of distortion (either at high or low signal volumes) can have similar perceived effects as restricting the frequency response. Unfortunately, there is no known way of testing the overall system distortion (including that of the photoelectric device), because it is difficult to make good test films.

8.7 Frequency responses

This section will extend concepts first introduced during the chapters on disc cutting, notably sections 6.4, 6.13 and 6.14, which I suggest you read first.

Both variabledensity and variablewidth soundtracks could be recorded by shining a beam of light through an electromechanical "light valve", comprising lightweight but stiff metal ribbons in a magnetic field. These ribbons were arranged to have a high frequency of resonance, usually between 7 and 12 kHz. Electrically, they functioned as resistors rather than inductors, so they gave constant amplitude characteristics on the film below these frequencies. That is to say, the resulting variations in width or density were substantially the same at all frequencies for constant inputs.

This also applies to variabledensity recordings made by exposing the film to a gaseous discharge lamp driven from the audio, or a modern photoemissive diode driven in a similar manner. In the vocabulary of Chapter 5, we always end up with a constant amplitude characteristic on the film. A great deal of high-frequency noise is masked by this technique, but the disadvantages are greater risk of highfrequency distortion, and worse lowfrequency hums and thuds.

Much of the work of the sound camera engineers consisted in minimising the high-frequency distortions, including volume limiters with side-chain pre-emphasis (Chapter 10) and innumerable intermodulationdistortion tests around each developing and printing stage. We must especially respect the wishes of those engineers today, especially since we are likely to add to the difficulties if we copy to another optical medium.

Linear noise reduction techniques like preemphasis were never used on releaseprints because of the desire to make things idiot-proof. But, within the studios, intermediate stages of the final soundtrack might carry a standard preemphasis curve remarkably similar to a modern digital preemphasis curve. Its time constants were 42 and 168 microseconds, giving a 12dB step in the recorded high frequency response between about 1000 and 4000Hz. However, it seems it was always used with the pushpull system to give greater fidelity in the stages prior to the final mix. You are unlikely to be in the business of extracting sound from such a component track; but if you are, it is worth knowing that (say) original music can be recovered with enhanced fidelity this way.

Another canofworms occurs when we are using cinema sound for another purpose (e.g. domestic video). In about 1940, the SMPE (Society of Motion Picture Engineers) introduced something called dialog equalization, to be applied to cinema film soundtracks to compensate for a number of psychoacoustic effects resulting from cinemas reproducing speech much louder than natural. (The thinking is explained very clearly in Ref. 5).

Therefore, it might well be advisable to reverse this standard (which I believe became universal) when converting cinema-film sound to another medium.

In 1944 the SMPTE Research Council proposed another standard equalisation characteristic to be used when 35mm films were printed to 16mm. This was designed to overcome the subjective woolliness of the severe highfrequency losses then inherent on 16mm optical soundtracks. A 1949 Revised Proposal flattened this equalisation somewhat, because there were less boomy loudspeakers and better H.F high frequency recording by then (Ref. 6), but at the same time dynamic compression was adopted. This latter was intended to overcome the fact that 16mm films were commonly shown with the noisy projector in the auditorium. None of these proposals became standards, but the principle remained - and still remains today - that a 16mm version of a 35mm film is very unlikely to have the same frequency and dynamic characteristics as the original.

8.8 Reducing background noise

During the days of the photo-electric cell, the reproducing equipment often contributed a significant amount of hiss of its own, even on 35mm film. It was necessary to make a trade off between having a broad optical gap to let a lot of light through, which meant concomitant equalisation, and a narrow gap which caused problems of hiss, focussing, and diffraction. The engineering journals of the time are filled with different people's compromises about this; but fortunately everyone seems to have recognised the idiot-proof advantages of inviolate constant-amplitude characteristics. All we have to do is follow in their footsteps with our improved technology.

Today, the background hiss of photoelectric devices is less of a problem, even on narrowgauge formats, but it still isn't good enough to allow much electronic extension of the high frequency response. All we can do is hurl as much light through the film as we can (provided we don't overload the electronics). We must focus it on the emulsion as sharply as we can (unfortunately this is often made difficult by idiot-proof designs), and keep the hum and noise of the exciterlamp as low as possible (hum is endemic on narrowgauge formats which tend to have A.C filament lamps). We could also explore the use of shortwavelength light in conjunction with new photoelectric devices to reduce the diffraction problems (there is room for new research here).

It seems scarcely worthwhile for me to say this, but the optical alignment of the sound system should start with a high-frequency loop film, and the maximum output should be sought. Next a frequency testfilm should be run, and the equalisation adjusted to get a flat result. Intermodulationdistortion testfilms exist for checking the uniformity of illumination and the azimuth of the slit, and buzztest films for centering the beam on the track; but both these parameters may require empirical adjustment when running actual films which have been printed imperfectly.

All the electronic techniques listed in Chapter 3 could be used to clean up the results, especially with Class A pushpull tracks or other systems which in effect give two versions of the same sound. We should probably not try to expand the dynamic range (the subject of Chapter 10), except when a narrowgauge version of a 35mm original is the only surviving copy; but the techniques of thump removal mentioned in chapter 10 are valuable for dealing with some tracks with imperfect noisereduction. As an outsider, I am also vaguely surprised that the equivalent of a liquid gate is not used to cut down some of the noise of scratches.

From 1974, release-prints on 35mm or larger may be encoded Dolby A or Dolby SR for added immunity from background noise when the track is split into two for stereo. Please see sections 9.4, 9.12 and 10.5 for further details.

REFERENCES

1: Van Urk, A. Th: "Sound Recorder of the Philips-Miller System," Philips Technical Review, 1936, no. 1, p. 135.

2: British Standard B.S.1404 "Screen luminance for the projection of 35mm film on matt screens" specified 8 to 16 foot-lamberts; and American Standards Z22.39-1944 and PH22.39-1953 "Screen Brightness for 35mm Motion Pictures" recommended 10 foot-lamberts plus or minus 1 foot-lambert. These recommendations are relaxed (i.e. the picture may be *dimmer*) off the axis of directional screens, or in outdoor theatres.

3: John G. Frayne and Halley Wolfe, "Elements of Sound Recording" (*book*), New York: John Wiley & Sons and London: Chapman & Hall, 1949. Chapters 15 to 20 deal with most of the listed types.

4: *ibid*., pp. 350-357.

5: D. P. Loye and K. F. Morgan (Electrical Research Products Inc.), "Sound Picture Recording and Reproducing Characteristics" (*paper*), Journal of the Society of Motion Picture Engineers, June 1939, page 631.

6: John G. Frayne and Halley Wolfe, "Elements of Sound Recording" (*book*), New York: John Wiley & Sons and London: Chapman & Hall, 1949, pp. 560-1.

9 Reciprocal noise reduction

9.1 Principles of noise reduction

Despite the hopes of early enthusiasts, magnetic and optical recordings did not eliminate background noise. Until then, the public had attributed such noise exclusively to the action of a needle in a groove. But magnetic recording also turned out to have a problem with background noise, due to the individual domains of unmagnetised tape each adding a random signal to the music, and we have already seen the effects of dirt upon optical recordings. Eventually noise reduction principles were applied to many analogue media, including VHS videocassettes, Laservision videodiscs, and some quadraphonic and stereo LP disc records.

If we measure the best analogue audiotape format available today (twin-track on half-inch tape at 30 inches per second), each track is capable of a signal-to-noise ratio of about 72 decibels. Each time the tape speed is halved, the signal-to-noise ratio suffers by three decibels (this isn't completely accurate because different recording characteristics have an effect, although the power-bandwidth product is definitely halved). And each time the track *width* is halved, the signal-to-noise ratio suffers another four or five decibels when you allow for the guard-bands. Thus, one track on the best available stereo ferric audiocassette will give a signal-to-noise ratio less than 50 decibels, and this is very noticeable.

I apologise for the fact that engineering vocabulary uses an ambiguous word, affecting this entire chapter. "Metre" is a unit for measuring "length," (in this case the width of a piece of tape), and "Meter" also means an artefact for measuring various phenomena (in this case, electrical voltage). I have called the latter device a "dial" to reduce the ambiguity, even though the device may not actually comprise a pointer moving across a scale.

9.2 Non-reciprocal and reciprocal noise-reduction

There are two ways of ameliorating the problem, known as "non-reciprocal", and "reciprocal" (or "complimentary") noise reduction systems.

Non-reciprocal systems are intended to be used on recordings which were made without any special processing, and rely on various psychoacoustic tricks to reduce the noise without apparently touching the wanted sounds. There have been a number of devices for this, ranging from the Scott Dynamic Noise Suppressor of 1947, through the Philips "DNL" (Dynamic Noise Limiter) of 1971, the "third stage" of the Packburn Noise Reducer of 1975, to currently-available devices made by Symmetrix and dbx, and the "hiss reduction" algorithm of CEDAR. You will find heated debate in the audio profession about the effects of these devices. There are two problems: (*a*) individuals have different psychoacoustic responses, so a machine which sounds perfectly satisfactory to one person will be creating all sorts of side-effects to another; (*b*) it is very difficult to balance the advantages of hiss-reduction against the disadvantages of no hiss reduction, because the original noise-free signal is not usually available for comparison, so it can only be a subjective judgement.

The author's view is that an archive has no business to be using them because of the principle I expounded in section 2.3. Although you should be aware that non-reciprocal noise reduction systems exist, and you may need them to *exploit* your archive's recordings, you should not use them for internal purposes.

But reciprocal noise reduction systems are a different matter. There are many of them. So far, they all work on the principle of boosting the volume of part or all of the audio before it is recorded, so it is stronger and drowns the background noise of the medium. Upon playback, the opposite treatment is applied, so the original sound is restored while the background noise is attenuated. Reciprocal noise reduction systems rely on psychoacoustics too, but the situation is slightly different, because the original sound is restored; only the background noise is modified. Psychoacoustics are involved only to mask the background-noise changing. It is still there if you analyse the reproduction, but it is certainly no worse than if you analysed the same sound recorded on the same medium *without* reciprocal noise-reduction. For the recording engineer, the choice then comes down to finding a system which works, is suited to the recording-medium and subject matter in question, and is affordable.

That last sentence implies that there are many solutions to the general problem of concealing background noise, which is true. I shall not waste your time on systems for *recording* audio, since my view is that all properly-engineered media using 16-bit linear PCM outperform analogue tape (in signal-to-noise ratio, if not in other parameters). But clearly the archivist may encounter any of the systems when playing other people's recordings.

I will therefore consider each system in approximately chronological order, describing its history and intended application, so you will be able to narrow down the possibilities if you have an undocumented recording. I shall also give any procedures necessary for recovering the wanted signal in its original state.

9.3 Recognising reciprocal noise reduction systems

We begin with some general remarks covering all the reciprocal noise reduction systems invented so far.

The first difficulty is to recognise *when* reciprocal noise reduction has been used. Listening to the raw signal can sometimes be confused with automatic volume controlling (Chapter 10).

Next we must recognise *which* system has been employed. In the absence of documentation this can be very perplexing; but since all reciprocal noise reduction systems modify the intended sound, we must neutralise them, both for objective and service copies.

The remainder of this chapter will give you a few clues to answer these questions, but there is no instant formula guaranteed to give the right answer. Fortunately, it is not quite like a subjective judgement of the sort which can give an infinite range of choices. Chronological evidence is the starting point, which often eliminates possibilities. Some systems require special alignment signals, which will provide a clue whenever they appear. There are only a finite number of systems, and most give results so dissimilar from others that it is usually a case of listening and making one quite unambiguous choice out of perhaps half-a-dozen possibilities.

Just occasionally, however, this doesn't work. For example, it is difficult to choose between "dbxI" and "dbxII," and between no noise reduction at all and "Dolby B." Also,

the "CX" system is specifically designed *not* to be audible. In these cases I can only recommend you do the job twice (documenting the difficulty), so later generations can pick the right one if they think they know better.

I am not an expert in writing software, but I am told that it isn't possible to emulate "Dolby B" (one of the simpler noise reduction systems) using digital signal processing, because the analogue version uses both impedance-mismatching and variable capacitances at the same time in order to get the desired transfer-characteristic. So it may be necessary to use analogue hardware for all Dolby B recordings, at least.

9.4 Principles of reciprocal noise reduction systems

The jargon phrase for putting an original signal through an analogue noise reduction processor before recording it is "encoding", and the phrase for dealing with it upon playback is "decoding."

Clearly the decoder must mirror precisely what the encoder did, or the original sound won't be restored. The two units must be tolerant of distortions inherent in the recording process, particularly phase-shifts. Most successful systems are designed with this in mind; but besides the ones I shall mention herewith, you may come across recordings made with prototype or unsuccessful systems of various provenances. Again, the ideal strategy is to do the job twice with explanatory documentation - once without decoding, so subsequent generations may have a go, and once to the best of your ability, so listeners can get your best simulation of the original sound in the meantime.

There are three basic ways of telling the decoder what the encoder did. You should understand these before you leap into action with all your different decoding devices.

The first method relies upon changes in the signal strength itself. For instance, if the signal going into the encoder rises by six decibels, the process may reduce this to three; this is known as "two-to-one compression." Then the decoder, receiving an increase of three decibels, knows that it has to expand this to six by a process of "one-to-two expansion."

The second method is similar, but limits the treatment to only part of the dynamic range to minimise side-effects; thus it is dependent upon the absolute signal strength. Although it is somewhat oversimplified, I can best express my meaning by a concrete example. The Dolby B system applies two-to-one compression to high-frequency signals at volumes 40 to 60 decibels below alignment level, therefore increasing these signals to only 40 to 50 decibels below alignment level. Louder signals remain untreated, so there can be no associated side-effects. Upon decoding, the unit takes no action unless the reproduced high frequencies are quieter than minus forty decibels; then one-to-two expansion takes place. The decoder must be set up correctly, so that the signals at minus forty decibels are presented to the decoder at exactly the same strength as they left the encoder. If this doesn't happen, the decoder may start to expand signals it wasn't meant to, or it may leave some compressed signals in their compressed state. Dolby Laboratories therefore specify a precise alignment procedure for the signal volumes, which we will consider later. If either the volumes or the equalisation are in error, you will not restore the original sound correctly.

The third method of controlling the decoder is, in effect, to send information about what the compressor is doing directly to the decoder. For sound recording, this means additional information must be recorded alongside the music. This method isn't used much, but was employed for the multi-channel optical soundtracks of the 1940 Walt

Disney film "*Fantasia*", where one of the soundtracks carried three line-up tones of different frequencies to control the expanders for three audio tracks. Some radio-mikes and lossless digital compression-systems use the same principle; although I'm not talking about those, you should know the possibility exists.

It should be noted that the first two methods of reciprocal noise reduction suffer from the disadvantage that any frequency response errors in the tape recording/reproducing process become exaggerated. Not only will the tonal balance of the original sounds be changed, but some rather complex things will happen to the dynamic properties of the original sounds as well. Ideally, all recordings made using these two methods should carry multi-frequency line-up tones, and in the case of the second method sensitivity-setting tones, to minimise the possible side-effects.

In order to control the volumes, both the encoder and the decoder must measure the volume of the appropriate part of the signal. Three quite different principles have been used for this as well. "Peak detection" relies upon the peak signal voltage, and is comparatively easy to implement. However it relies upon the circuitry (and the analogue recording medium) to have an infinitely fast response. Sometimes there are advantages in having a peak signal voltage detector which is deliberately slowed down, and this is the second principle. Although it's a misnomer, this modification is known as "average" signal detection, and the response in milliseconds should be quantified for correct results. (It never is!) "R.M.S detection" (Root-mean-square detection) is the third principle. It measures the amount of *power* in the signal. It takes an appreciable amount of time to get the answer, because at least one cycle of audio must be analysed, and typical RMS detectors for audio may be tens of milliseconds "late." In addition, the circuitry is complex. But this method is inherently resistant to phase-changes, which often occur with analogue tape-recorders. On the other hand, peak detection is suitable for low-level signals, because it is found that quieter sounds usually have fewer transients and comprise slowly-decaying waveforms for which peak detection is reasonably consistent. I mention all this because you should understand the principles if ever you have to reverse-engineer a recording for which you do not have a decoder.

There is another difficulty with controlling the decoder from the wanted signal. Inaudible ultrasonic frequencies (such as stereo radio pilot-tones, or television line-scan frequency interference) may "block" an encoder. Such tones are generally too high-pitched to be recorded accurately, so mistracking will occur on playback. To solve the first problem, many noise reduction units or cassette-recorders are fitted with multiplex filters (often labelled "MPX ON"). These will eliminate the 19kHz pilot-tone of stereo radio broadcasts before the encoder. But television interference has always been a problem, because it occurs between 15kHz and 16kHz, which are audible frequencies for many people, and should not be filtered off. Decoding such tapes may require some creative engineering to emulate the original fault.

The next eight sections describe different reciprocal noise reduction systems.

9.5 Dolby "A"

History

The earliest reciprocal noise reduction system to achieve commercial success was "Dolby A", first used at the English Decca studios in April 1966. History tells us the first record made with the system was Mahler's 2nd Symphony (Decca SET325-6). In 1967 the

system became generally available, and was soon adopted for professional multitrack studio work. This was because mixing several separate analogue tracks greatly increases the basic hiss level, so hiss becomes particularly noticeable on the final recording. Each channel cost several hundred pounds. It was a year or two before Dolby A was employed for much straight stereo work, but after that Dolby A remained the dominant noise reduction system in top professional applications until Dolby SR began to displace it in 1987. It was used by the film industry for internal post-production purposes at an early date, but the Model 364 for cinema projection was not made available until February 1972, after which the optical soundtracks on release-prints were often coded Dolby A. It may sometimes be found on C-format videotapes from 1980 onwards. But Dolby's licensing scheme specifically prevented it from being used in domestic applications, as did the price!

Method of Operation

The system divides the frequency range into four bands so that a psychoacoustic effect called "masking" prevents the rising and falling tape-hiss in one band from being audible in another. When encoding, each band has two-to-one compression giving ten decibels of noise-suppression, except the highest frequency band where the effect amounts to fifteen decibels. Only low-level signals are treated. Dr. Ray Dolby was very concerned that his pioneering system should not produce any audible side-effects, and basically he succeeded; only one or two electronically-generated signals from synthesisers have been known to "catch it out."

Line-up Procedure

The first version of the unit was intended to sit across the inputs and outputs of a perfectly-aligned Decca tape recorder working to NAB characteristics (see section 7.8). The mixing-console was supposed to provide the necessary alignment-tones and metering. Later Dolby units had dials of their own for measuring the tones, and when the Model 360 range appeared in 1970 a line-up tone generator was built into each one.

In order to keep all Dolby A tapes compatible (so they could be edited together, for example), Dolby Laboratories insisted that contemporary Decca line-up procedures be followed. The NAB specification not only deals with reproduction characteristics (Section 7.8), but also the levels of recorded signals, and it was arranged that all Dolby A tapes should be provided with a section of NAB line-up tone (which, to be quantitative, was a magnetic strength of 185 nanowebers per meter). (Ref. 1). When played back, if this read to the "NAB" mark on the Dolby unit's dial, correct restoration of the sound was assured. Unfortunately, other European users used a different standard measurement ("DIN"), corresponding to 320 nanowebers per meter, about four and a half decibels higher. This was used by EMI and the BBC, and corresponded to the peak recommended signal volume, since the best tapes of the time had two percent total harmonic distortion at this point. So Dolby units were supplied with another mark on their dials corresponding to DIN volume; but the Model 360s only generated tones at the NAB volume, and in practice this has now become the standard for aligning Dolby units.

A Dolby-level test tape is therefore needed (this applies to Dolby B and Dolby C as well). Because practical tape-heads may have a small degree of misalignment, they may not play separate tracks correctly, but will read some of the unrecorded "guard-band" between tracks, giving the wrong answer. A Dolby test tape is therefore recorded across

the full width of the tape. Even on a misaligned cassette-recorder, this ensures the machine will record at the right strength and play its own tapes back correctly, although head realignment may be necessary to get the optimum output on another reproducer as we saw in Chapter 6.

Each Model 360 was designed to generate a special kind of tone so it could be recognised anywhere. It was pitched at around 400Hz, but had a warble about once per second. To the uninitiated, it sounds as if the tape is catching on the flange of the spool! But it's intentional, and provides almost certain evidence that the recording which follows is coded Dolby A. The decoding Dolby must normally be set up so this tone reads to the "NAB" mark on the dial.

Of course, you may be exceedingly unlucky and find a Dolby A tape without such a tone - the earliest Dolbys did not have such a tone generator, as we've seen - and the only advice I can offer is to listen to undecoded Dolby A tapes whenever you can, so you will recognise the characteristic sound of such a tape if it should arrive out of the blue. (It's not unknown for Dolby A film soundtracks to be broadcast in their undecoded state, for example. Also British Telecom sometimes used Dolby A to reduce the noise of its analogue landlines, occasionally forgetting to switch in the decoder.)

9.6 Dolby "B"

History

The Dolby B system was introduced specifically for domestic use. Instead of making the units themselves, the company licensed other firms to manufacture the circuit, thereby encouraging it to be built into machines for convenience. The licensing scheme specifically forbade its use on 15ips machines so there could be no confusion with Dolby A. Between early 1968 and 1970 the system was exclusively licensed to American tape-recorder manufacturer KLH, who included the circuit in their Model 40 tape recorder. A year later other makes of open-reel machines, and stand-alone units such as the Advent and the Kellar, were being made with Dolby B circuitry. From 1979 quite a few VHS video recorders had the system built-in for stereo linear soundtracks. But the greatest application was for the medium of the audiocassette, which was just becoming popular when the first cassette machines were made with it in mid-1970. It is no exaggeration to say that Dolby B changed the audiocassette format from a curiosity into a fully-fledged quality recording system for amateurs, and the company encouraged this by asking for no royalties on pre-recorded cassettes made with the system.

Method of Operation

When recording, Dolby B compresses high frequencies by 2: 1 between uncoded levels of -40 to -60dBs. Whereas Dolby A divides the frequency range into four bands, Dolby B divides the range into two, and processes only the high frequency one where the tape hiss is most noticeable. In domestic situations hiss is the main problem, while professionals are worried about other noise all the way down the frequency spectrum (such as hum, print-through, and magnetised heads).

The company *did* insist that pre-recorded cassettes be unambiguously labelled with the Dolby trademark if they were coded Dolby B, but the author knows many which weren't. Because much of the spectrum is untreated, it can sometimes be difficult to

decide whether Dolby B has been used on an undocumented tape. Amateurs seem very bad at documenting these things, and I regret I've found professionals no better, apparently because they think an audiocassette cannot be a serious format. Again, I recommend you to get familiar with the sound of a Dolby B recording. Although it affects only the high frequencies, the effect is in some ways more conspicuous than Dolby A, because the balance between low and high frequencies is altered. On the other hand, misaligned machines can be weak in treble because of cheap cassettes or faults such as azimuth errors or overbiassing. So it can often be difficult to disentangle all the reasons. Fortunately the situation is helped by most cassette players having a DOLBY ON/OFF switch, so it is relatively simple to try it and see.

Line-up Procedure

Early versions had line-up generators and dials like Dolby A, but the tone was generally a steady frequency in the neighbourhood of 300-400Hz with no "warble."

With the advent of standard cassette formulations (see section 7.9) this was found largely unnecessary. But, for the pedantic, Dolby B line-up tone is 185 nanowebers per meter on open-reel tape, and 200 nanowebers per meter on cassettes, and alignment of the circuit should be done using a test tape as described in the section on Dolby A.

Since standardised formulations came into use, it is distinctly unusual for there to be any line-up tone. If the machine is misaligned or the audiocassette is not of standard sensitivity, "Dolby mistracking" can occur, even when you have correctly identified whether the tape is Dolbyed or not. The clue is usually that high-frequency sounds, such as sibilants of speech, seem to fall into a hole or sit on a plateau with respect to the low-frequency sounds. Assuming your machine has been correctly aligned (you will need a Dolby level test tape, a service-manual, and a dial for this), the fault must lie with the cassette or the alignment of the original machine.

In these cases, the best solution is a separate free-standing Dolby unit, such as the JVC NR-50 (which also offers two other systems as well). Rather than de-aligning the innards of your cassette player, you can switch its Dolby off, and twiddle the volume controls on the inputs and outputs of the free-standing decoder until you've minimised the problem. But this is an empirical process, and for archival reasons you should ideally make an undecoded version as well.

9.7 DBX systems

General History

After Dolby, the next significant noise reduction system was that due to dbx Inc. I should explain there was an early stage before it was marketed as a *reciprocal* noise reduction system. Although the company did not call it such, I call this early implementation "dbx 0." The first implementation which was specifically a *reciprocal* noise reduction system was called "dbx Professional," but by 1974 there were compatible units with phono connectors and other features to appeal to down-market users (the 150 series), so this came to be known by everyone as "dbx I" for short, and this is now official. The second has always been called "dbx II." I shall consider all three systems here.

Method of Operation

The basic idea is to compress the dynamic range of sounds before being recorded, and expand them again upon playback. Thus, the original sounds would be restored, and the background noise of the recording medium drowned. The difficulty has always been to create a circuit in which degrees of compression and expansion could be reliably controlled over a wide dynamic range. The breakthrough came in 1970 when David Blackmer patented an integrated-circuit configuration in which amplification could be controlled extremely accurately over a range of a hundred decibels or more. (Refs. 2, 3).

Line-up Procedure

None needed - all dBx systems operate equally at all practical signal levels.

The "dbx 0"

Blackmer's integrated-circuit was first marketed in a consumer unit for compressing or expanding musical signals subjectively. There were actually two models, known as the Model 117 and the Model 119, with slightly different facilities. The dominant feature of both was a large control knob which varied the compression or expansion with infinite resolution. In one direction, you could select compression ratios from one-to-one to infinity-to-one; in the other direction the box would function as an expander from one-to-one to one-to-two.

The new integrated-circuit actually did this very well. The intended application was to expand analogue sound media such as broadcasts or LPs to give a full dynamic range, rather than the manually compressed versions then available for consumers. So it did not alter the frequency balance, and for some time it was the only system in which the encoded frequency-balance was the same as the uncoded version.

Although it was not specifically marketed for its noise reduction abilities, hi-fi buffs soon learnt that it would perform this function quite well. By dialling up (say) 1.5: 1 when recording, and 1: 1.5 on replay, the original dynamic range could be restored. And if a value in the neighbourhood of 1.5: 1 was chosen, one could also keep one's options open and hear compressed or expanded music by suitable operation of the knob when replaying. It was particularly useful in video work, because television pictures are commonly watched under less-than-ideal listening-conditions, and the "dbx 0" permitted a compressed signal for such purposes without attacking the frequency balance or committing the recordist to irrevocable compression. (This is explained in Chapter 10, sections 11.6 onwards; the "irrevocable" element is explained in section 11.11).

But it had side-effects. There was no attempt to mask the effect of tape hiss going up and down (it wasn't made with this in mind, of course); and to maintain stereo balance, both channels went up and down together. But it has a place in history as a *de facto* noise reduction format simply because there wasn't anything else like it at the time.

The "dbx I"

In 1972 David Blackmer marketed a version specifically for tape noise reduction purposes (Ref. 4). This differed from "dbx 0" in several ways. There was now a separate processor for each channel, and the compression was fixed at 2: 1 (so the expansion was 1: 2). By a technique called "sidechain pre-emphasis" and careful selection of the timing parameters of the operation, the effect of tape-hiss going up and down was largely masked. As the

compression was effective across the whole dynamic range, the signal-to-noise ratio of any tape recorder was effectively doubled, although Blackmer conservatively put the improvement at thirty decibels; this was much more than Dolby could offer. The system also circumvented the problems of Dolby alignment-tones. So it had a number of appealing features, especially for smaller studios which could not afford top-grade equipment (or the time to align it).

"dbx I" is still marketed and used for professional applications. As far as I am aware, there are only three criticisms. (*a*) When used on musical instruments comprising high-pitched notes without any bass content, you can sometimes hear the *low*-pitched noise of the tape going up and down. (*b*) If the recording and playback sides of the two analogue machines do not have exact responses at low frequencies, barely-audible low-frequency errors can be magnified into large dynamic errors. (*c*) And, inherently, *any* frequency response errors appear to be doubled upon playback, although not when the signal comprises a wide range of frequencies all present at the same time.

The "dbx II"

The second and third criticisms in the previous paragraph are mitigated (but not cured) by "dbx II", introduced about 1976. The solution adopted is to deliberately restrict the frequency-range going into the control circuitry so that frequency response errors have less effect. For semi-professional applications this is a great help, although the protection against sources of noise at either extreme of the frequency range suffers in consequence. Dbx II is therefore recommended for such applications as amateur audiotapes, amateur and professional films and videos, and audiocassettes, where alignment of the tape recorder to give a perfectly flat response across the entire frequency range isn't practicable.

Notes on the use of dbx Equipment

A disadvantage of many types of dbxII decoders is that they have a capacitative input. You are therefore warned not to use them to decode a recording played on a machine with a high-impedance output (more than about 2000 ohms), or the high frequencies will be attenuated before the decoder, which will make things worse. (Ref. 5)

For a year or two in the early 1980s a few LP disc records were made in America with dbx II coding. The Model 21 decoder was made especially for discs (it wouldn't encode), and it was offered for as little as £20 to encourage its use for records. There were at least nine record companies making dbx-coded discs (including Turnabout, Unicorn, Chalfont, and Varese Sarabande); but the system was not used by bigger record companies, so the repertoire did not have many significant artists. And, in the author's experience, the system was impotent against the real bugbear of disc records - the loud click. When clicks were the same loudness as the music the expander was unable to separate them, and when they were even louder the expander actually made them worse.

The "hi-fi tracks" of VHS video recorders are encoded with dbx II to overcome the head-switching noise (section 7.15). Most of the time the system works, but you can easily hear the side-effects if you try recording a pure low-frequency or high-frequency tone on the hi-fi track of the video. You do not notice the effect on consistently loud wideband audio, which comprises most TV sound these days. The circuit is "hard-wired" in this application. To make things simpler for video users, there is no option to remove it.

In the absence of documentation, detecting a recording made using one of the three dbx systems really needs someone who is familiar with the *sound* of an encoded version (as usual). There are no line-up tones to provide a clue. The "compressed sound" of all three dbx systems is very apparent: inoffensive background noises such as mike hiss or distant traffic can be magnified by alarming amounts during gaps. But dbx 0 can be separated from the others very easily because the tonal balance of the wanted sound remains the same. For dbx I and dbx II, the encoded signal has extra "presence" or brightness.

As I mentioned earlier, it is sometimes difficult to distinguish between dbx I and dbx II recordings in the absence of exact documentation; this may not be anybody's fault, because there *was* no "dbx II" for a couple of years. But you can assume dbx I is confined to professional applications; I have only found it on multitrack and quarter-inch open-reel tapes. So long as the wanted sound has a restricted frequency range, the two systems have the same result, so they become compatible anyway. The problem only occurs on sounds with a wider range. I am afraid the only way you can decide is to listen to low-level low-frequency signals decoded by each system in turn, and choose the one with fewer side-effects.

9.8 JVC ANRS (Audio Noise Reduction System)

This was an attempt by the Japanese Victor Company to make a noise reduction system to rival Dolby B without infringing Dolby's patents. It was introduced in 1973 for JVC's audiocassette machines. Whether there was a deliberate policy of creating something similar to Dolby B, or whether it was a case of convergent evolution, I do not know. JVC claimed "Dolby B music tapes can be played back through ANRS," which is obviously perfectly true, literally speaking! In my limited experience Dolby B circuits seem to decode ANRS recordings quite satisfactorily. The same procedures for signal volume alignment apply.

A different version of the ANRS circuit was incorporated in JVC's "CD-4" quadraphonic coding system for LP discs (see section 10.6).

In 1976 JVC attempted to "gild the lily" by extending the ANRS circuit for cassette tapes. This version was called "super ANRS", and recorders with this facility were downwards-compatible because they included standard ANRS as well. The super-ANRS attempted to compress loud high-frequency sounds of the type which caused overloads in the days before chrome and metal tapes and the Dolby HX-Pro circuit (section 9.14). It gave between six and twelve decibels more headroom, depending how good the basic machine was. However it was very vulnerable to HF response errors, and even when these were right, its effect was accompanied by very noticeable "breathing."

By 1978 JVC realised that it was Dolby's trademark which was encouraging the sale of cassette recorders, not theirs, so they quietly retired their own version and became official Dolby licencees.

9.9 Telcom C4

This was a noise reduction system specifically invented to combine the advantages of Dolby A and dBxI. So, from that, you will gather it was aimed at professionals. It was introduced by Telefunken in Germany in 1976. It has remained the dominant professional

noise reduction system in Germany and Scandinavia, where it is still used on professional linear videotape soundtracks as well; it has become the *de facto* standard on "B-format" machines, which are also German in origin. Only two British commercial sound studios used it (The Angel in London and Castle Sound in Edinburgh), although many more have Telcom "c4 DM" cards which plug into Dolby frames to provide Telcom on an existing tape machine without hassles.

The unit divides the frequency range into four bands, rather like Dolby A, but it processes each band by compressing at 1.5 to 1 on record and expanding at 1 to 1.5 on playback. The crossover frequencies are 215Hz, 1.45kHz and 4.8kHz. There are no "thresholds", so there is no need to align the sensitivity. It is claimed the relatively gentle slope of 1.5 to 1 gives a sufficient degree of noise reduction while minimising the multiplication of response-errors in the tape recorder. Perhaps you should also be aware that there is another version (model 112S) intended for analogue satellite links, which has a slope of 2.5 to 1.

9.10 Telefunken "High-Com"

This was a version of the Telcom C4 introduced about 1983 for down-market use, mainly by small studios. I regret I know nothing about how it worked. The circuit was marketed by Rebis of Great Britain and by D&R Electronics of Holland as one of the possible options for their uniform rack-mounted signal-processing units, also targeted at the emerging "home studio" market. Another version, Hi-Com II, was made as a free-standing unit by Nakamichi, presumably for the top-of-the-range cassette market.

9.11 The "CX" systems

Originally this was a system invented by the American record company CBS for reducing surface-noise on conventional LP records. The acronym stood for "Compatible Expansion," and the idea was that records made with the system would be partially compatible with ordinary records, so that users without CX decoders wouldn't notice the difference. Therefore the frequency-balance of the encoded version was kept the same as the uncoded version, and CBS engineers devised a rather complicated system of decay-times operating at different levels which minimised the side-effects of the gain going up and down. In my opinion they were successful at this.

Unfortunately, the basic philosophy was never made clear by CBS executives at the system's launch in 1981. It was never intended that the compression should be inaudible, only that the side-effects should be minimal. The philosophy was that volume compression of various types - both manual and automatic - were then normal on most LPs anyway (as we shall see in chapter 10). The CX system was designed to replace these techniques with something which could be reversed with consistent results. Compression was happening already; it was decided to make a virtue out of the necessity.

Unhappily, not only did the sales people claim that CX records were fully compatible (when they were never meant to be), but the press demonstrations were handled badly which alienated serious journalists. (Refs, 6, 7). In the writer's view this was a great pity. The CX system offered the first and only way in which bridges could be built between sound systems with limited dynamic range. It would, for example, have been possible to introduce the system in a large organisation like the BBC, which was never

able to adopt reciprocal noise reduction on its internal recordings, because of the muddles which would occur when recordings made on differing dates in different studios were transmitted. CX would have permitted wider dynamic range when required for post-production purposes, while the effect would be barely audible if an encoded tape was accidentally broadcast without decoding. Thus the only chance of creating a real upgrade in technical standards, for use whenever recordings could *never* all be made to the same standards at the same time, was thrown away.

Unhappily again, the engineering staff at CBS were evidently commanded to make various modifications to the "standard" in an attempt to do the impossible - make a noise reduction system which you truly couldn't hear. In particular, they reduced the compression for when hissy master-tapes were encoded (resulting in excessively hissy LPs). Basically this was the fault of the sales staff again, who were trying to dress mutton as lamb without listening to the results. For the ghastly details, see Reference 8.

So far as I am aware, CX was only used on pre-recorded media. As with Dolby, different compression-ratios took place at different levels, so nowadays we have to go through a level calibration procedure when playing back. On LPs the calibration tone was 3.54 cm/sec RMS for each channel alone, which should just trigger a LED on the back of the decoder. A 7-inch stereo calibration LP (catalogue number CBS CX-REF) was provided with each decoder.

Now to details of where you may come across CX-coded recordings. In America the system was used between 1981 and 1983 on LPs made by CBS and Gasparo (and I gather it was marked on the sleeve using a very small trademark - so look closely), and on Laservision and CED videodiscs. In Europe CX was never used on LPs, only on some Laservision videodiscs (and these to a different "standard"). Unfortunately CX decoders are not common in Europe, and the various "standards" mentioned in Reference 8 mean that (sooner or later) you will have to attack a decoder with a soldering iron anyway. It will obviously be quicker to see if the material has been re-released on compact disc first (it usually will be)! But in other cases, it does seem the results would mean the labour is worthwhile.

9.12 Dolby "C"

By 1981, it was becoming apparent that the highly-successful Dolby B system was being overtaken by events. The intervening decade had brought startling improvements to the performance of the audiocassette format, and dBx had shown there was a demand for greater dynamic range. Ray Dolby therefore introduced his "Type C" system. This was licensed in such a way to ensure Dolby B could be encoded and decoded as well.

To oversimplify slightly, Dolby C comprised two Dolby B-type circuits in series. The new half gave a further ten decibels of noise-reduction. It operated half as fast, at signal levels ten decibels lower, and on lower parts of the sound spectrum, so the two halves could not "fight" each other. The result was twenty decibels of noise reduction at most frequencies above 1000Hz, falling to five decibels as low as 100Hz. There are other considerations in the design (Ref. 9), but that's the basic idea.

Dr. Ray Dolby cunningly licensed his scheme to cassette recorder makers at no extra charge, thereby in effect prolonging the Dolby B patent. The system was included in most manufacturers' top-of-the-range cassette machines after about 1983, but very few pre-recorded audiocassettes were made with the system.

Careful level alignment is necessary, as with Dolby B; the procedure is the same as Dolby B, since the Dolby C circuit contains the essence of Dolby B as well. The system seems to work best on cassette-machines of the type which have "self-alignment" programs (for adjusting their bias, equalisation, and recording-levels to suit the cassette in question).

Dolby C was included on the linear stereo soundtracks of one VHS recorder (the Panasonic N7300), and is used for the linear soundtracks of the professional Betacam and Betacam SP video systems. (Dolby B is not allowed on these because licencees are not allowed to use Dolby B for professional applications. If you don't like Dolby C on your Betacam, your only option is no noise reduction at all).

Unhappily, the author has found Dolby C not to be as perfect as Dolby B. Certainly when the circuit works well, it does so with no audible side-effects, and makes a great reduction in perceived hiss. But when tapes are played from machines which did not record them, distortions of various types are sometimes heard, and Dolby C seems more vulnerable than Dolby B. I suspect, although I have no proof, that because peak detection is used (see section 9.3), anomalies can occur in the new section of electronics which covers the mid-frequencies, where these effects are more noticeable. I find I can clearly hear it on the majority of television news stories shot on Betacam SP, although the result is certainly better than no noise reduction at all. Fortunately, the free-standing JVC NR-50 can be useful for decoding rogue Dolby C recordings as well.

9.13 Dolby SR and Dolby S

"SR" stands for "Spectral Recording." This was a new noise reduction system for professional applications introduced by Dolby in late 1986. The circuit gives about 26 decibels of noise reduction with (apparently) no audible side-effects. Its launch stunted the sale of professional digital audio recorders, because it was much cheaper to convert existing analogue machines to Dolby SR. In addition, many people claimed its sound was better than digital, whilst the technology was also familiar to users.

The new feature was that, instead of fixed crossovers to allow "masking", the crossovers slid up and down the spectrum depending upon the psychoacoustic effects measured in "barks" (section 4.18). The Dolby Alignment Tones hitherto needed for aligning the A, B and C systems were replaced by a system of "pink noise" fifteen decibels below "Dolby level." This sounds like a hissing noise similar to the inter-station noise of an FM radio receiver, but quieter. Although this hiss must be recorded at about the right sensitivity and frequency-response to ensure editing can always occur, the decoding Dolby SR unit switches between the reproduced hiss and its own internally-generated hiss, so if any differences are heard the operator can realign his tape-machine. The pink-noise is interrupted every couple of seconds by a twenty-millisecond gap. Thus it can be assumed that any tapes preceded by such a noise must be encoded Dolby SR.

Dolby SR is now being used for optical film soundtracks. Although Dolby A films give acceptable results on non-Dolby projection equipment, this is not true of SR. The problem is rather more "gain-pumping" than can be tolerated; the actual frequency-range is over-clear, but balanced. Dolby SR sometimes appears to be used to "improve" 16-bit digital recording machines, since compact discs coded SR sometimes "escape".

Dolby have announced a simpler version for domestic use, called "Dolby S". Its first appearance was on narrow-track multitrack tape recorders for the home studio market, where its main rival was dBxII; Dolby S is much more "transparent," and from its

principles I should expect it to be the best system in this application. It is now available on many up-market cassette machines.

Some of BMG's pre-recorded cassettes have Dolby S coding. The system designers say there is a degree of compatibility with "Dolby B." (Dolby Labs do not have a reputation for making remarks like that, unless they're true). It was probably stimulated by the fact that Digital Compact Cassette (DCC) claimed "compatibility", since DCC machines would *play* analogue cassettes; but they could not *record* them!

9.14 Other noise reduction systems

The following reciprocal noise reduction systems have been marketed at various times, but have not achieved much success. It will therefore be difficult for you to recover the original sound if you encounter recordings made with them and you do not own a suitable decoder. So I append what little information I know, so an enthusiastic operator may simulate them, or an enthusiastic circuit-designer may imitate them.

ACCESSIT Compander. This British company marketed a noise reduction system in the mid-1970s which compressed 2 to 1 on record and 1 to 2 on replay. All frequencies were treated equally, so you would think it would be compatible with "dBx 0"; but unfortunately this knowledge is not sufficient to ensure correct decoding, because I do not know whether RMS or peak detection was used, nor do I know the recovery-times.

ACES. A British-made system "with 2 to 1 compression/expansion ratio", as usual, available from about 1984.

BEL BC3. This was a British-made system introduced in 1982 aiming at the market occupied by "dBx II", that is home studio recordists and small multitrack studios. Although it is clear that true compatibility was never intended, I gather it makes a pretty good job of decoding dBxII encoded material.

BNR. This is short for "Beta Noise Reduction," and was provided for the linear soundtracks of Sony Betamax videocassettes before "Hi-Fi audio" was recorded by the picture head-drum. It was supplied on the first Betamax recorders with stereo sound in 1982, because the already narrow audio track was split into two even-narrower ones. In the writer's experience, it failed to conceal the hiss going up and down, and demonstrates that noise reduction always works best when there is no noise to reduce! Measurements with meters show 2:1 compression; but what is going on in the way of pre-emphasis and sidechain pre-emphasis isn't clear. A Sony Betamax video machine with the appropriate circuitry will be needed to play it; as far as I know, the only two models marketed in Britain were the Sony C9 and the Sony SLO1700.

BURWEN "Noise Eliminator" Model 2000. This was introduced by Burwen Laboratories of Burlington, Massachusets, in 1971, and was easily the most expensive reciprocal noise reduction system ever sold, at over £3450 per stereo unit at a time when there were 2.4 dollars to the pound. It was very powerful and made very impressive noises at demonstrations; it compressed at 3 to 1 and expanded at 1 to 3 for most of the dynamic range. (It was linear at low levels). Thus a professional analogue tape recorder would have a theoretical dynamic range exceeding 110dB. Each unit offered three "characteristics."

Characteristic "A" was optimised for 15ips tape, characteristic "B" for 7.5ips tape, and characteristic "C" was intended for lower tape speeds, discs, and FM broadcasting. (This latter characteristic ganged the two stereo channels together, but was otherwise like "B".).

One freakish feature incorporated bass equalisation between 10 and 25Hz on *record* to overcome *replay* losses at low frequencies. This was done to ensure correct processing at low frequencies, always difficult on full-range reciprocal noise reduction systems. But there were two fundamental objections to the principle: (*a*) the NAB standard was violated, and (*b*) there was no way the tape could be correctly restored on another machine in the absence of a set of clearly-documented low-frequency line-up tones. Whatever the reason, the seminal review in Studio Sound for February/March 1974 (in which the three professional systems Dolby A, dbxI, and Burwen, were compared using laboratory measurements and operational tests) showed it could not restore the dynamic range of an original sound correctly. That review killed it.

DO-IT-YOURSELF SYSTEMS. Various circuits for do-it-yourself electronics enthusiasts have been published. Most compress at 2 to 1 and expand at 1 to 2. A famous one by Reg Williamson had straightforward pre-emphasis rather than pre-emphasis in the side-chain, and average detection. (Ref. 10). Another by Dr. David Ellis had no pre-emphasis at all, but used average detection which sped up at lower signal levels (Ref. 11). But the reason for the popularity of home-made devices can be understood from the latter's claim that four channels of simultaneous encode/decode could be built for as little as £50.

SANYO "SUPER D" (Model N55). This attempted to overcome the low-frequency "pumping" of dBxI (section 9.6) by splitting the frequency range into two (the crossover being 2kHz), and compressing each half separately at 2: 1. Since high frequency sounds are, in general, lower volume than low frequency sounds, the resulting tape sounds toppy, and can be confused with dBx. Unfortunately, practical tape recorders generated intermodulation products behind high frequency sounds, which the low-frequency expander brought up on playback. This gave the curious effect that intermodulation distortion sounded *worse* as the recorded volume went *down*! To reduce saturation difficulties, high frequencies above 8kHz were compressed even more strongly; a 1kHz oscillator provided an alignment-tone for this.

TOSHIBA ADRES (Automatic dynamic range expansion system). ADRES stand-alone adaptors, and the ADRES system incorporated into Toshiba cassette recorders, were marketed in the years 1981 and 1982. Reference 12 described the system as having a compression ratio just under 1.5 to 1, with varying pre-emphasis according to input level, giving about 22 decibels of noise-reduction. Like Dolby, alignment-tones were used. When this writer briefly tested a unit, the line-up tone frequency was observed to be 1kHz with no warble. But I was unable to find what the recorded level was supposed to be; it could well have been the same as Dolby-level. Obviously the need for line-up tone implies the unit behaved differently at different levels, but simple measurements and listening-tests suggested the unit behaved like two different systems in series. First, a section with a consistent compression ratio of 1.41 to 1 (giving +7dB out for +10dB in) with a flat frequency response and a relatively slow recovery-time. Second, a Dolby-B style top lift at lower levels, starting with +1dB at 6kHz for inputs 20dB below line-up level, and with a very rapid recovery-time. The unit had no audible side-effects, but failed to conceal low-pitched tape noise.

9.15 Noise reduction systems not needing treatment

As a footnote to this chapter, I shall tell you about two methods of sound recording which have been called "noise reduction" systems, but which are not reciprocal processes, so they need no treatment upon playback.

Optical Noise-Reduction is a method for reducing the background noise of optical film soundtracks. It was first used in 1932 and is universal today. (See chapter 7, Fig. 9.1E). Films of the 1930s were sometimes specifically advertised as having "noise reduction." Do not be confused by this; no action is needed to restore the original sound upon playback.

Dolby "HX-Pro" was a system originally invented by Dolby Laboratories and known simply as "HX" (for "Headroom Extension") in June 1979. The idea was to reduce the A.C. bias of a magnetic recorder when high frequencies were being recorded, to reduce the self-erasure effect. (See section 7.3). It was found that the wanted audio could shake up the magnetic domains in exactly the same way that bias did, so by juggling with the bias in conjunction with the audio, more intense high frequencies could be put onto the tape without distortion. Thus the power-bandwidth product of the tape was increased.

The original HX circuit got its information about the presence of high-frequency signals from the Dolby B encoder; but it was soon realised this was something of a shotgun wedding, and instead Dolby did more development in conjunction with tape recorder manufacturer Bang and Olufsen in Denmark. The result, called "HX-Pro", was independent of any reciprocal noise reduction system and meant better recording of high-frequencies under any conditions. It was provided on quality cassette machines from late 1981 onwards, and since then some open-reel tape-recorders have used it. At least one pre-recorded cassette publisher (Valley of the Sun Publishing) has not only packed chrome tape in cassette shells designed for ferric, but printed its inlay cards with the Dolby trademark plus microscopic letters "HX PRO"; but again, no action is needed upon playback.

Also I should mention "Dolby E". This is not a form of reciprocal noise reduction at all, but a lossy digital compression system (section 3.6), permitting up to eight channels of digital audio to be carried along two AES standard digital cables, together with Dolby Surround metadata (section 10.13). It also has the advantage that, unlike AC-3, audio frames match video frames, so Dolby Surround may be handled in a video environment.

9.16 Conclusion

I should like to end this chapter by repeating my warning. You must constantly be on the lookout for undocumented and misapplied noise reduction systems. In the six months since I started writing this chapter, I personally have come across a broadcast which was coded Dolby SR, a VHS video-soundtrack coded BNR, a quarter-inch tape *de*coded Dolby A instead of *en*coded Dolby A, and a commercial compact disc coded dbxI. Note that none of these was documented (let alone expected). Also note they weren't just my personal opinions; subsequent research (usually by laborious comparison with other versions of the same subject matter) unambiguously *confirmed* my diagnosis every time.

This is one of the areas in which you can prevent future generations from either calling us clots, or totally misunderstanding analogue recording techniques. Keep *your* ears open!

REFERENCES

1. The exact calibration of magnetic strength upon recoded media is a very complex matter, and the two standards authorities on either side of the Atlantic have different test-methods which give different results. The ANSI standard (American National Standards Institute) is now known to give results about 10% lower (about 0.8dB) than the DIN method (Germany). Thus the ANSI (NAB) test-tone of 185nWb/m measures about 200nWb/m when measured by the DIN method, while the 320nWb/m of the DIN standard measures about 290nWb/m by the ANSI method. In each case, I have quoted magnetic reference-levels measured by the standard method for the appropriate continent - i.e. ANSI for NAB tapes - even though Dolby A was developed for Decca's NAB machines in Europe!

2: U. S. Patent 3681618.

3: Ben Duncan: "VCAs Investigated - Part 2," *Studio Sound*, Volume 31 No. 7 (July 1989), page 58.

4: David E. Blackmer, "A Wide Dynamic Range Noise Reduction System," *db Magazine* Volume 6 No. 3, August-September 1972, page 54.

5: Peter Mitchell, "The dbx Z Problem," *Hi-Fi News*, January 1982, pages 37-39.

6: "Business", by Barry Fox. *Studio Sound*, November 1981, page 98.

7: "Business", by Barry Fox. *Studio Sound*, December 1981, page 56.

8: "CX - an approach to disc noise reduction," by John Roberts. *Studio Sound*, March 1983 pp. 52-53.

9: "A 20dB Audio Noise Reduction System for Consumer Applications", Ray Dolby; *Journal of the Audio Engineering Society*, Vol. 31 No. 3, March 1983, pp. 98-113.

10: Reg Williams (*sic* - should have been Reg Williamson), *Hi-Fi News*, May 1979 pages 84-87, and June 1989 pages 56-63.

11: "Noise Reduction Unit" by Dr. David Ellis, *Electronics & Music Maker* (a magazine published by Maplin electronic component retailers, exact date unknown), reprinted in "The Best of Electronics & Music Maker Projects Volume 1", Maplin, 1983.

12: Mike Jones, "Review of Aurex PC X88AD", *Hi-Fi News*, November 1981, pages 112-113.

10 Spatial recordings

10.1 Introduction

This chapter considers the special problems of reproducing and copying sound recordings which have a "spatial" element - stereophonic, quadraphonic, and so on.

In case you didn't know, I should make it clear that no perfect system of directional sound reproduction will ever be achieved. The plain engineering reason is that (for loudspeaker listening, at least) the amount of information we need to reproduce is many orders of magnitude greater than we need for high-definition TV, let alone audio. But there are more subtle reasons as well. Human beings use many sound localisation mechanisms, which are learnt and correlated in early childhood. Different individuals give different weights to the various factors, and the sense of hearing works in conjunction with other senses (especially for this purpose the senses of sight, balance, and the sensation our muscles give us as they move). And there are other factors you wouldn't think would make a difference (such as the relative importances of reverberation in the recording and playback environments, the degree of curvature of sound waves, etc.).

No recording can contain all the information needed for accurate spatial reproduction. In the meantime, recording engineers have made compromises, managed performances, and used undocumented "trade secrets" to achieve their ends. So this chapter mainly helps you *reproduce* the *intended* original sound, not restore the original sound faithfully. It will guide you through the processes of conserving what directional information there is, nothing more. For an important but brief summary of the problem, I urge you to read Ref. 1.

Spatial recordings generally comprise two or more "channels" of audio-frequency information. "Channels" is rather a woolly term, because in some circumstances one channel can be split into two or two can be combined into one; so I shall use it to mean "intended separate audio signals." This isn't much better; but it does allow me to present you with the options in order of increasing complexity.

10.2 "Two-channel" recordings

By this I mean recordings which consist of two separate sound recordings made in synchronism - two "channels." The first difficulty occurs when the same subject matter appears both in single-channel and two-channel versions. You might go for the two-channel one, on the grounds that inherently it has a better power-bandwidth product; and ninety-five percent of the time, you'd be right. But always remember that occasionally this isn't true. Usually, a stereo recording can be turned into a mono version by paralleling the two channels. You should remember that sounds appearing equally on both channels will appear on the mono version three decibels higher than sounds which appear on one channel alone. Therefore you should think carefully before abandoning the mono "equivalent" without checking it meets this specification. (A consistent exception to the rule will be encountered in section 10.4 below).

Another difficulty might be that the single-channel version has a better power-bandwidth product simply because it was done on a better recording machine, or because

alternative restoration processes can be used (such as the one described in paragraph 7 of Box 4.17 – see section 4.17), or because the two versions have been edited differently. You'd expect me to recommend the two versions be treated like two separate recordings in these cases; but the rest of the time you can assume mono listeners will simply parallel the channels if they want to, and you need only do the job once.

Nowadays, the word "stereophonic" has settled down to mean a two-channel recording intended to be replayed on a matched pair of loudspeakers in front of the listener. But this has not always been the case; you should be aware that not all recordings conform to this definition. For example, the words "stereo" and "stereophonic" were used pretty indiscriminately before the mid-1950s to denote a single-channel recording which had a particularly clear "sense of space" around the performers. You will find the words in publicity-material and record reviews, and on the labels of at least one series of commercial mono discs (the Columbia "Stereo Seven" series of 7-inch 33rpm LPs of 1951-2, published in the USA). All these items are single-channel mono, and should be treated accordingly.

To confuse matters further, in 1960 the British Standards Institution advocated that the word "Stereophonic" should be printed on all disc record labels where the groove contained a vertical element. This was done with the best of intentions, to prevent consumers from damaging such records by playing them with lateral pickups; but it does not necessarily mean that the sound is "true stereo" (recorded with two separate channels of sound). Many such records are "fake-stereo" in some way. This is not the place to explore all the ways in which a single-channel recording can be turned into a fake-stereo one, but the techniques include frequency-division into two channels, inserting 90-degree phase-shifts between the two channels, manual panning, and adding two-channel reverberation to a single-channel original. To the restoration operator, the only question is "how to restore the original sound." My opinion is that if you can't get access to a single-channel original, you should transfer the two channels as they are, since that's obviously what the second engineer intended; but document it as a fake where you know it for a fact, and leave the users to decide what to do about it (if anything)!

Sometimes the "fake-stereo" version may have a better power-bandwidth product than the original mono one. We must then neutralise the fake-stereo effect in order to follow the principles I mentioned in section 2.3. This will require the operator to discover which fake-stereo process was used, and reverse-engineer it where possible.

This leads us to the topic of "compatible stereo" records. Given a correctly-adjusted stereo pickup, most stereo disc records are "compatible" in the sense that satisfactory mono reproduction occurs when the two channels are mixed together. But a few discs were made with deliberately-reduced vertical modulation so they would not be damaged by a lateral pickup. This was done by restricting the amplitude of the "A minus B" signal - the *difference* between the two channels. This was helped because the human ear is less able to distinguish the direction of low-frequency sounds. The effect might be achieved acoustically (by using two directional microphones whose polar diagrams degenerated to omnidirectional at low frequencies), or electronically (by introducing a bass-cut and/or an automatic volume limiter into the difference signal).

Frequently, this type of compatibility would be achieved subliminally. In a multi-mike balance, for example, most engineers panned bass instruments to the centre, which resulted in lateral vibrations of the cutter (analogous to a mono disc). Certain resonances in a church or hall might be recognised by a balance-engineer as causing lots of (A-B), so he might re-position the microphone to make the effect sound similar for a mono listener, as well as making easily-playable stereo discs. And I know at least one record - for which I

was responsible! - where my difference-signal was so great, the record company reduced my tape to mono and turned it into a "fake stereo" one.

In principle, at least, it is possible to reverse-engineer some of these artefacts, and restore "true stereo" when transferring the sound to a medium which is *not* an LP disc. But *should* this be done? My personal answer to this is "No - but document it!" In other words, we should not manipulate the (A-B) signal when doing the transfer, but to document when the original disc claimed to be "compatible stereo," so users may perform the operation on the copy if they wish. And research whether there has been an unmodified re-issue on CD!

Another terminological difficulty concerns the words "stereo" and "binaural." Nowadays, the former implies that the recording is meant to be heard on loudspeakers, and the latter on headphones. (We shall meet another meaning of "binaural" in section 10.5). Different microphone techniques are normally used for the different listening conditions, and it is the duty of the copying operator to ensure the labelling is transferred to the copy so that future listeners take the appropriate action. But in the 1950s the two words were often used interchangeably, and further investigation may be necessary to document how the sound was meant to be heard. Fortunately, only the documentation is affected; the transfer process will be the same in either case.

In 1993 Thorn EMI's Central Research Labs introduced "Sensaura Audio Reality," a process which seeks to modify separate tracks of sound in the digital domain to produce suitable effects for headphone listeners. The published descriptions are (perhaps deliberately) ambiguous about whether the process is meant to sound better on loudspeakers or on headphones (Refs. 2 and 3). We shall have to listen to examples of the same music processed in two different ways before we can decide; but the point remains that you should document the intended means of playback.

"Dummy-head" stereo is the principal form of "binaural" stereo (in the sense of headphone listening). The recordings are generally made using two omnidirectional microphones spaced about six inches apart with a baffle between them. The baffle might have the same shape and size as a human head. For obvious reasons, the technique of having a human head on top of a pole doesn't often happen; much binaural recording is done with a flat transparent baffle between the microphones.

In the remainder of this section, I shall outline some two-channel systems which are supposed to be better than ordinary stereo or binaural ones. None of them will require special treatment when archive or service copies are made, although nearly all will require transfers to an uncommonly high standard. But the documentation should carry the name of the system, so users will know what to do with the transfer. So we start with some terminological matters.

"Holophonic" recordings are also intended for headphone listening. The published descriptions of the invention were surrounded by an enormous amount of hocus-pocus, and it is difficult to establish exactly what was involved (Ref. 4); but a conservative interpretation might be that a sophisticated "dummy-head" was used. I understand that the acoustic properties of human flesh and the detailed structure of the outer ear were simulated to help resolve ambiguities between front and rear images. Whatever the hocus-pocus may have meant, most people found it an improvement over conventional dummy-head stereo; but it proved remarkably difficult to maintain this improvement through mass-production processes, which shows how important it is that conversion to digital must be done to extraordinarily high standards, if at all.

We revert to loudspeaker listening for our remaining examples of two-channel terminology. "Ambisonic Stereo" is a special case of "Ambisonics," a spatial-sound

processing technique which has much wider applications than just stereo. Section 10.11 will give a fuller description.

The "Hafler System" takes the difference between two loudspeaker stereo signals and sends it to a third loudspeaker (or pair of loudspeakers) behind the listener's seat. Certain recordings made using coincident bidirectional microphones contain out-of-phase information which represents sounds picked up from the sides of the array. The Hafler circuit takes this information and puts it round the back. It distorts the geometry of the original location, but it reproduces genuine sound from outside the "stereo stage." The circuit can be applied to many stereo recordings, and some were marketed specifically for this treatment.

"Q-Sound" (Ref. 5 and Ref. 6) and "Roland Space Sound" (Ref. 7) are systems for loudspeaker listening dating from the early 1990s. They claim to take advantage of some psychoacoustic effects to permit reproduction rather wider than the "stereo stage."

"Dolby stereo" isn't really a stereo system at all; we shall leave it till section 10.8. Here again I must repeat an earlier point. Various digital compression systems which sound perfectly satisfactory on plain stereo have been known to give surprising side-effects when applied to Dolby stereo soundtracks. So strict rigour in the digitisation process is essential. The spatial information is often very vulnerable, and must be conserved.

10.3 Archaeological stereo

In recent years, some workers have attempted to get "stereo" from two single-channel recordings of the same performance made through different microphones, the two recordings not having been synchronised in any way. This might happen when the same concert was being recorded commercially as well as being broadcast, or when a record company insured against breakdowns by recording with two independent sets of kit. The first documented example seems to have been Toscanini's 1939 NBC broadcast of Copland's *El Salon Mexico* (Ref. 8), where NBC had rigged microphones for networks in the United States, and another with Spanish announcements for NBC's affiliates in South America. The difficulties are (*a*) identifying sessions done this way, and (*b*) synchronising the two recordings with sufficient accuracy (ideally a few microseconds).

One might also ask *whether* this process should be used, since it does not represent the wishes of the contemporary artists or engineers. In any case, the two microphones are extremely unlikely to have been positioned correctly for either stereo or binaural purposes. My view is that so long as the two single-channel versions are not corrupted by the synchronisation process, and there are no other side-effects of creating the "stereo version," there is no harm in trying to get closer to the original sound; but obviously the archive copies must be left alone.

I should dearly love to be able to tell you about restoring such sounds, but personally I have never been able to get beyond stage (*a*) in the previous paragraph but one. The technique is simple enough, though, as described by its inventor Brad Kay (Ref. 9). The recordings are approximately synchronised, equalised, and run at almost the same speed and volume. Sooner or later they will crawl into synchronisation and out again. As they do, the effect is heard upon a pair of loudspeakers. What usually happens is that a double-mono effect occurs, which can be verified by switching with one of the items phase-reversed. If this gives a characteristic "phasing" effect, then it's two records of the same signal; but apparently this doesn't happen every time. The next stage would be to

achieve synchronism which gives a stereo effect; but having never got this far, I cannot predict what should happen next. Personally, I would put the two recordings on a digital editor with "varispeed", manipulate one so it synchronises approximately with the other, and reproduce the result through a Cedar "azimuth corrector" which is programmed to improve stereo using certain psychoacoustic algorithms (end of section 7.6). And clearly that must be confined to "service copies" only.

In section 2.6 I spoke about the duties of a professional discographer trained to locate several copies of a given recording. We will now have to ask him to check if they make stereo pairs. For 78rpm discs, visual inspection is the first sign. A few discographical clues exist (not enough to make any hard-and-fast rules). For example, recordings made by branches of the European "Gramophone Company" marked the two matrixes differently. Duplicates of takes were marked with an A after the take-number, and this practice was continued after the formation of EMI in 1931. The presence of an A means "look out for plain takes as well." But most of the time it seems the two machines recorded the same electrical signal, and there is no stereo to be had. Also when EMI started using tape recorders (about 1948), it became the practice to give a "plain take" to a direct-cut disc and an "A" to a tape-to-disc version of the same performance, so the idea is unlikely to work post-1948. If two versions of the same take were published, it is likely that one would be used in Europe and the other overseas (e.g. in America), so the discographer's job isn't going to be easy.

This is the place to record that the HMV "Artist's Sheets" on microfilm at the British Library Sound Archive show coded details of engineering parameters for the years 1925-1931. All the evidence supports the theory that Columns 2 and 3 contain serial numbers of the cutting-amplifier and microphone respectively. If the "straight takes" and the "-A takes" have one piece of equipment in common, then the records must be "double-mono." For pseudo-stereo, we *must* have pairs of recordings for which *both* serial numbers are different.

The English Decca Company is thought to have run its "ffrr" and non-"ffrr" cutters in parallel between 1941 and about 1945, calling the former "Take 2" and the latter "Take 1." Decca did not document its matrixes very well, and did not export many, so only a test-pressing from those years will be likely to lead to success.

And in the 1930s there were numerous minor makes where the same material was recorded on discs of different diameters, or on conventional 78s and "microgroove" versions. These might all bear investigation (but I haven't found anything myself).

As we saw in chapter 6, stereo recordings only work well when there is pretty tight synchronisation between the two channels. You should be aware that some digital recording systems (for example, the "EIAJ" system - used for the Sony PCM-F1 encoder) do *not* record the two channels synchronously. They economise by using use *one* analogue-to-digital converter, switched between the two channels at an ultrasonic rate. This doesn't matter so long as the same process is used on replay - the sound comes out synchronously then - but if you are doing any digital processing in the meantime, or converting to another system (e.g. the Sony 1610) in the digital domain, you will have to take this into account. The time-difference between two channels sampled at 44.1kHz is eleven microseconds. This makes a barely-detectable shift in the stereo image-location (to the left), and causes a slight loss of treble (again barely-detectable) when the channels are paralleled for mono.

To make matters worse, one manufacturer of digital interface boxes did not know whether it was the left or the right channel which came first, and got it wrong in his

design. Thus it is absolutely vital to check that centre-mono sources come out in synchronism from every digital recording, irrespective of the provenance!

10.4 Matrixing into two channels

Two-channel recordings also exist which are not always meant for reproduction on *two* loudspeakers or earpieces. Sounds may be "encoded" from three or four audio-frequency channels, and distributed on two audio-frequency channels, in anticipation that they will be "decoded" back into three or four loudspeakers. This principle is known as "matrixing," and so far all the systems which have been commercially marketed are supposed to give compatible stereo results. They too can be stored in the form of two-channel "archive copies," but we shall consider how they may be decoded (e.g. for "service copies") in section 10.8.

For the time being, that is all I have to say about two-channel recordings. In the majority of cases, all we need to do is transfer them as well as we can, documenting any unusual features as we go. But when we come to systems with more than two channels, we may be obliged to do considerable research to discover *how* the sounds were meant to be heard. The remainder of this chapter basically comprises such evidence.

10.5 Three-channel recordings.

First, another piece of terminology. When a four-channel recording system starts off as four different channels (for example, four microphones), which are encoded into two channels of audio bandwidth, and then decoded back into four to be reproduced on four loudspeakers, the system is known as a 4-2-4 system. When a system comprises four original sound channels which are kept separate (or "discrete") through all the distribution processes and end up on four separate loudspeakers, the result is known as a 4-4-4 system. This convention can be extended to most forms of spatial reproduction, and from now on I shall be forced to use it, starting with *three*-channel recordings.

It is a commonplace of the film industry that everything is stored in threes - "music," "effects," and "dialog." But I shall not be dealing with this type of three-channel recording here; I shall be dealing with three-channel versions of the *same* sounds. Such recordings are more common in America than in Europe. They are confined to magnetic media, usually half-inch tape or 35mm sepmag film. They usually occurred for one of three reasons, which I shall now outline.

The first reason was to make a two-channel stereo recording using two spaced microphones, plus a third mono recording with a microphone midway between the two spaced microphones. This enabled stereo and mono versions to be accommodated on the same tape and edited in the same operation. It reduced the artefacts of phase-cancellation for the mono listeners. Also, if three tracks of equal dimensions were recorded, it gave each of the stereo tracks the same hiss-level as the mono track. This idea counts as a 2-2-2 system plus a 1-1-1 system on the same tape.

Pairs of spaced stereo mikes often gave faulty stereo imaging on playback - the notorious "hole-in-the-middle" effect. This could be reduced by using the third track for genuine three-channel stereo, which automatically filled the "hole-in-the-middle." Three-channel stereo was favoured in cinema applications, because if there were only two loudspeakers situated at each side of a wide screen, any member of the audience sitting

substantially off the centre-line would perceive the stereo image collapse into the nearer loudspeaker. A third loudspeaker mitigated this. This counts as a "3-3-3" system. The Warner "3-D" system (used for *The House of Wax* in 1953) had a 35mm sepmag soundtrack made to this philosophy, but it is comparatively rare to find it on complete final-mix film soundtracks.

For domestic listeners, who could be assumed to be sitting in the central "stereo seat" in preference to paying for a third channel, the third microphone could be used to create a stable phantom image between the stereo loudspeakers by using conventional panpot techniques at the studio. This was usually essential in (for example) a violin concerto. With only two spaced microphones picking up the orchestra satisfactorily, a soloist in front of the orchestra might disappear into the "hole-in-the-middle," or any small movements he made might be ridiculously exaggerated as he moved closer to one of the two spaced mikes. These problems could be ameliorated by rigging a third spaced mike; but it was risky to combine it electrically and record the concerto on only two channels. Careful judgement was needed to balance the soloist (who was very close to the middle microphone) against the orchestra (which was not), and achieve a satisfactory compromise for both stereo and mono listeners. This could be done more satisfactorily after the three-track tape was edited and before it was transferred to the master disc. This counts as a 3-3-2 system; and the resulting two-channel tape was often called "binaural" to distinguish it from the *three*-channel one.

Thus we have three separate reasons for the existence of three-track stereo tapes. You will not be surprised to hear that I consider all three tracks should be transferred onto the "archive copy" using a multitrack digital medium; but for a satisfactory service copy, we need to delve into the *reasons* for the three-track tape. If a contemporary two-channel version exists, we shall have to use the three-channel original for its power-bandwidth product, and compare it with the two-channel version to establish how the centre track was used. Was it used only for the mono version? Or was it was used throughout the stereo version (and if so, at what level)? Or was it mixed "dynamically"? (We should certainly need to make a special "service copy" in the latter case, for it would be impracticable to remix it every time someone wanted to hear it!) Another difficulty arises if it had been a genuine three-channel stereo recording in the cinema, and we can only provide two channels for the listeners to our "service copy." We may need to remix the three tracks into two using a Dolby Stereo Encoder to approach the original effect. And - document it!

10.6 Four-channel recordings - in the cinema

First, some details of four-channel systems used in the cinema.

The first major three-dimensional picture, *The House of Wax* (1953), used two picture projectors for the left and right eyes, plus a third three-track 35mm sepmag film carrying sounds for three loudspeakers behind the screen, as we saw in section 10.5. There was, however, a fourth channel on one of the picture reels; it was intended for effects from behind the audience. I have no information about how well the fourth track worked, but I have my suspicions! In any case, Warners foresaw the difficulties that such a complex projection system would involve, and arranged for the optical soundtrack of the other picture to have a straightforward mono track. Most cinemas either showed the film with "one-eyed" mono or "two-eyed" mono from this comopt track. Since it wasn't a wide-screen film, the disadvantages weren't great.

Shortly afterwards, "CinemaScope" films were introduced specifically to compete with television, using a wide screen. The first manifestation in Britain was in November 1953, using three-track sepmag like *The House of Wax*. But Twentieth-Century Fox soon found a way of dispensing with the sepmag machine. Four magnetic stripes were attached to the 35mm film, which was modified by providing smaller sprocket-holes to allow the necessary space. Three of the tracks were 0.063 inches wide and gave reasonable quality; these were used for the loudspeakers behind the screen. The fourth track was intended for occasional effects behind the audience. It was only 0.029 inch wide, so it was rather hissy. When the fourth track was meant to be silent, a 12kHz gating-pulse was recorded, so the rear loudspeakers were muted instead of emitting hiss. (Ref. 10).

10.7 Four-channel audio-only principles

Now we return to sound-only media, and (I'm afraid) some more terminology.

First, "quadraphony" (there are variations in the spelling of this). This generally means a system for domestic listening in which four original signals (e.g. four microphones) are reproduced on four similar loudspeakers - preferably in a square - surrounding the listener in the horizontal plane. Proceeding round the listener clockwise, the individual loudspeakers are usually "left front," "right front," "right back," and "left back." Experienced listeners criticised this idea because the front speakers would then subtend an angle of ninety degrees at the listener, instead of the sixty degrees generally considered optimum for stereo; but I won't bother with the various compromises designed to circumvent this difficulty. Instead, I shall concentrate on pure square reproduction whenever I use the word "quadraphonic."

"Surround" (used as a noun or an adjective) usually implies a system which might be for domestic or auditorium reproduction in conjunction with pictures. It comprises a two-channel soundtrack which gives reasonable reproduction when played back on a pair of stereo loudspeakers to a listener in the ideal stereo seat. But a "surround-sound" soundtrack can be extended to give extra channels - for example, one midway between the stereo loudspeakers (to "anchor" dialogue, as with "Dolby Stereo" which we shall consider later), and one behind the listener (this may not be used continually, but for occasional special effects). A few audio-only versions have appeared on compact discs, but they never seem to work as well without the pictures, underlining that we need to remember there is interaction between our senses of hearing and of sight.

Few subjects have caused more confusion than the subject of "matrixing," especially in the early days when it was used for squashing four channels into two for Quadraphonic Sound. No-one was able to agree on the *aims* of the process, let alone its basic implementation. Were the encoding and decoding matrixes supposed to be "transparent" to an ideal discrete four-channel master-tape, or was the master-tape supposed to have been pre-configured with the deficiencies of the subsequent matrix in mind? Was the aim to give perfect downwards-compatibility with stereo, or give degraded stereo at the expense of better multi-channel sound? Did the matrix have to cope with four loudspeakers that were not identical? or not in a square? Was the priority to give accurate reproduction in the front quadrant, or the front and sides (ignoring the back), or what? Were all the signals to be kept in correct phase relationships in (*a*) the encoded and (*b*) decoded versions, and if not, what compromises were acceptable? What

about mono compatibility? or height information? or listeners off-centre? or listeners sitting on typists chairs who could turn and face any direction? And so on.

The confusion was made worse by the fact that no matrix system was perfect, and different manufacturers provided different "directional enhancement" techniques to improve their results. These often meant that a decoder made for System A's records and actually used on System B's records might sound better than System B's decoder. I shall ignore these enhancement techniques; in an archival context, they constitute "subjective" interference. But if you're interested in seeing how they worked, please see Ref. 11. The basic theory of a "linear matrix" (i.e. one which works the same at all signal volumes) is complex but fairly well known (see Ref. 12).

I shall now consider the principles behind the major systems used for pre-recorded Quadraphonic media.

10.8 Matrix "quadraphonic" systems

The 4-2-4 quadraphonic systems and the 4-4-4 quadraphonic systems require different approaches from the archival point of view. We shall start with 4-2-4 systems, i.e. ones requiring a "matrix."

Two 4-2-4 systems were promoted in the early 1970s ("QS" and "SQ"), and ("Dolby Stereo") in 1976. All three systems claimed to be "downwards-compatible" so far as the stereo listener was concerned. That is to say, if the listener didn't have a decoder, he would just get acceptable stereo, making a "4-2-2 system". None of these systems was "mono-compatible;" there are always problems with sounds intended for reproduction on the rear pair of loudspeakers. So a fourth was developed by the BBC for broadcasting purposes, where both stereo and mono compatibility were important. ("Matrix H").

An "archive copy" of a 4-2-4 recording needs only to be on two channels, since expansion back to four can be done at a later date. That's the point of principle; I do not actually advocate the use of a decoder for making a four-channel "archive copy." But I appreciate an institution may rebel against installing decoders whenever such a recording has to be played, and therefore I could understand its desire to make four-channel "service copies."

So, although it isn't essential for the purposes of this manual, I shall say something about how 4-2-4 matrix decoding may be achieved, so you will do justice to what the engineers intended. I shan't give all the technical details, I shall only describe the outlines in words. If you seek mathematical definitions of the various matrixes, please see Ref. 12.

10.9 The QS system

This was a development of "the Regular Matrix" invented by Scheiber (Refs. 13 and 14). In Britain about 400 "QS" records were published in the years 1972-4, mostly by the Pye group, and the system was used by the BBC Transcription Service (notably for the wedding of Princess Anne). Unfortunately, BBC policy prevented these recordings from being advertised with the name of the commercial process, so I cannot advise you which ones used it!

I shall first describe the "Regular Matrix," because it was supplied on a great many domestic amplifiers, although as far as I know there were no commercial records

published with it. A point-source sound would be processed by the encoder as follows. If centre-front, it would be sent equally on the two transmission channels, just like centre-front stereo. All the remaining angles from the centre-front line would be halved. In other words, a left-front quadraphonic signal would be treated as if it were half-left stereo, centre-left quadraphonic would become left-front stereo, and back-left quadraphonic would become left-front with an out-of-phase component. Thus all the sounds picked up quadraphonically would be reproduced on a pair of stereo loudspeakers, although not necessarily in phase. All sounds intended to be in the rear two quadrants for the quadraphonic listener would have an out-of-phase component, and would decrease in a mono mix. When decoded back onto four loudspeakers, a theoretical point-source resulted in output from at least two (usually three) loudspeakers. (In other words, there would be a large amount of crosstalk - actually -3dB - between neighbouring channels, although the crosstalk was perfect between diagonal channels).

The "QS" system comprised the above matrix with a modification proposed by Sansui to mitigate the crosstalk disadvantage. (Ref. 15). Before being matrixed, the two rear signals were subjected to phase-shifts of +90 and -90 degrees with reference to the front channels; QS decoders had circuits after the matrix to shift the rear outputs by -90 and +90 degrees to restore the status quo. In the meantime, some types of crosstalk were slightly reduced, at the cost of ninety-degree phase-shifts in the reproduction of sounds in the side quadrants.

In practice, two techniques were adopted to ameliorate the remaining crosstalk. The first depended on the fact that a "perfect quadraphonic microphone" didn't exist. No four-channel microphone was ever made - or *could* be made - which would pick up sound from *only* one quadrant of the horizontal plane and give *only* one output. However, if the operator in the sound control room listened to the signals through an encoder and decoder, it was possible to juggle with practical microphones empirically to get improved results.

The other technique was incorporated in Sansui's QS decoder. It recognised what the engineers were trying to do, and it adjusted the gains of the four outputs to enhance the desired directionality. It was a proprietary circuit taking advantage of some psychoacoustic precedence effects; it did not claim to "restore the original sound." Despite this, the circuit was very successful on single instruments or small groups, but gave anomalous results on simultaneous sounds from all round the circle. In fact, it was Sansui's circuit which made the system attractive to the critical ears of the BBC Transcription Service.

The QS system cannot be reverse-engineered to give the directional characteristics of the original; it can only give you what the engineers were hearing in the control room. Whether you use a Sansui QS Decoder for a better "service copy" or "objective copy" is up to you, but it has no part in making the "archive copy," which must be the encoded two-channel recording.

10.10 The SQ system

This was also a "4-2-4 quadraphonic" system as defined above. It was invented by the makers of CBS (Columbia Broadcasting System) Records in America, and was used by the EMI group in Britain between 1972 and 1980. There were possibly a thousand issues in all. Unfortunately, Columbia was an EMI trademark in Britain, so CBS records were not allowed to be imported with that trade-name. The British market was insufficient to

warrant printing new sleeves and labels, so quadraphonic CBS records were "grey imports" in Britain. Although a large catalogue of imports was offered, the discs themselves appeared with stickers over the sleeve and label logos.

Unlike QS, SQ left the left-front and right-front signals alone, so if the wanted sounds were confined to the front quadrant, SQ discs could also be sold to stereo customers and played on two loudspeakers without any perceptible difference. (The system altered the actual reverberation, however, and EMI only marketed the two versions as "compatible" from October 1975 onwards).

For full details of how the coders and decoders worked, please see Ref. 12 and Ref. 16. Briefly, the two rear channels were encoded as follows. A left-back quadraphonic signal was split between the left stereo and right stereo channels with 90 degrees of phase shift. A stereo disc cutter would respond to this by tracing a clockwise spiral as viewed from the front of the stylus. A right-back quadraphonic signal would have -90 degrees of phase shift, resulting in an anticlockwise spiral path. Thus all the sounds picked up in the front quadrant would be reproduced correctly on a pair of stereo loudspeakers, but sounds from other quadrants would be reproduced on a stereo system with phase-reversed components. So long as such sounds comprised natural reverberation this wasn't too serious. But for single-point sounds positioned at the sides or back of the quadraphonic image, sound-mixers had to listen to the effect through a coder/decoder system and make their own judgements to obtain a satisfactory sound. Sounds intended to be directly behind the quadraphonic listener would disappear in mono, so CBS advised recording engineers not to locate soloists in that position.

A system called the "Tate DES" (Directional Enhancement System) was applied to SQ decoders in 1982, but counts as a "subjective" system for our purposes. The Tate DES was later adapted for "Dolby Surround."

The SQ system cannot be reverse-engineered to give the directional characteristics of all the original sound; it can only give you what the engineers were hearing in the control room. Your archive copy must be a two-channel version. Your objective copy must be a four-channel transfer through an SQ decoder. (Reference 17 gives details of a "do-it-yourself" SQ decoder. The performance of a SQ decoder may be checked with test disc CBS SQT1100). Your service copy may either be this or an undecoded two-channel transfer, depending on whether the subject matter was confined to the front quadrant or not; the implication of EMI's announcement of October 1975 is that SQ recordings issued before that date will need to be on a four-channel medium.

10.11 Matrix H

This was a matrix system invented by the BBC in 1975-6 with the specific goal of being equally compatible for stereo and mono listeners. Since mono listeners were then in the majority, this was a laudable aim.

When the final reports by the BBC Research Department were completed, it was apparent that there had been several versions. In 1975 the "old boy network" repeatedly carried rumours that secret test transmissions were being contemplated. This was to see if the compatibility claim was met before the final version was announced. But in fact the first transmission was during a 1976 Promenade Concert, and after that one or two concerts were advertised in advance as being "Matrix H."

The Hi-Fi Press castigated these for their "phasiness." That is, there were complaints that stereo listeners were being fed an undesirable amount of phase-shift

between the left-front and right-front channels (on top of that expected from ordinary stereo reproduction). It isn't clear which version was being used for which transmission, but versions with inherent 55-degree and 35-degree shifts were tried.

With the collapse of the commercial quadraphonic market in 1977, "Matrix H" was quietly buried. But off-air stereo recordings of that period will have some quadraphonic information in them. It is debatable whether the engineers of the time would prefer us to decode them into four channels now! The last version of the Matrix H decoder (Ref. 18) was never marketed commercially, but relied on the QS Directional Enhancement process. When tested with skilled listeners, the processes gave only 77% success compared with discrete quadraphony.

10.12 "Dolby Stereo"

Some history and some more terminology. Many early "wide-screen" films were on 70mm stock, which incorporated seven magnetic soundtracks. There were five intended for reproduction behind the screen and two "effects" (surround) channels - a "7-7-7" system. This was a pictures-only medium; but the next development became used in a sound-only context.

In 1974 Dolby Labs decided it was time to upgrade conventional (35mm) cinema sound, and they decided from Day One that they needed only three channels behind the screen and one "surround" channel. To get these four channels onto 35mm stock Dolby introduced a new type of matrix circuit. It was called (rather misleadingly) "Dolby Stereo", although it was actually a 4-2-4 system, not a 2-2-2 system; for this reason, I shall always put the phrase between inverted commas.

Two optical channels on the release-print were located side-by-side and encoded with the Dolby A noise reduction system (section 9.4); but a cinema not equipped for "Dolby Stereo" could get acceptable results using mono projectors which played both channels, when it functioned as a 4-2-1 system. Simplified versions of the matrix were marketed for the public from about 1984, when the first pre-recorded videos using "Dolby Stereo" were issued to the public. The first domestic version was called "Dolby Surround," and it had a linear matrix. This was upgraded to a "non-linear" one (with additional direction-enhancing circuits) and marketed with the name "Pro-Logic" in 1986; this, to all intents and purposes, duplicated what the professional "Dolby Stereo" matrixes were doing ten years earlier. A Pro-Logic decoder can also be used as a 4-2-3 system (dispensing with the rear channel) to improve television stereo sound; this was given the tradename "Dolby 3 stereo," but nobody seems to use this name nowadays.

Instead of a square array of loudspeakers (as with Quadraphony), the four loudspeakers are intended to be in the geometrical shape known as a "kite." (Definition: A kite is a plane figure bounded by four straight lines. Two of the lines which meet at a certain vertex are equal in length, and the two lines which meet at the opposite vertex are also equal in length). Generally, the system is meant to accompany pictures, so I shall assume this point from now on; but a few audio compact discs have been made using "Dolby Stereo" to take advantage of the user-base of Dolby Surround and Pro-Logic matrixes.

Ordinary stereo reproduction comes from loudspeakers L and R, while the viewer sits somewhere near the centre of the kite at V[3]. A picture-display (whose width is denoted by AB) is placed between the stereo loudspeakers L and R. In the cinema AB is

[3] An ilustration for this and the next 3 paragraphs is missing.

nearly as long as LR, but in domestic situations it is not usually practicable to get a video screen big enough for the angle LVR to be optimum for stereo (about sixty degrees). So LR may be four times the size of AB.

A central loudspeaker at C is often advantageous, and cinemas always use one behind the centre of the screen; but any reasonable-sized domestic loudspeaker will be obstructed by the video screen, so it will have to be fitted beneath. The purpose of such a centre loudspeaker is to "anchor" dialogue for off-axis viewers, so the voices appear to come from the picture. This centre loudspeaker only needs to be capable of handling speech frequencies, so good bass response isn't vital. Therefore another option is to place *two* (paralleled) good-quality "bookshelf-sized" loudspeakers immediately adjacent to the video screen at A and B (provided their magnets don't affect the picture). However not every domestic viewer likes this degree of complexity, and most Dolby Surround and Pro-Logic decoders have an option for no centre-loudspeaker at all.

The rear loudspeaker is at R, and again this might actually be more than just one. (Cinemas commonly have three or five, plus more if there's an upper circle). In practice, film soundtrack personnel do not route sound continuously to the rear, but use it for special dramatic effects. Such dramatic effects usually involve loud and powerful low frequencies (e.g. gunfire, aircraft, bombs, etc), so the rear speaker(s) must be able to handle lots of low-frequency power. It is also helpful if they do not sound like *point* sources of sound, and sometimes placing them at varying heights helps.

The rear-channel sound is actually encoded as out-of-phase information in the two main channels (so it will disappear if the two main channels are reproduced in mono). I therefore repeat my message that no matrix system can conserve all the directional information; you *cannot* have a surround-sound system *and* retain downwards compatibility at the same time, and this is why "Dolby Stereo" licencees are generally accompanied by a Dolby Laboratories representative during the mixing sessions. Various tricks can be invoked at the post-production stage to ensure the listeners believe they're hearing the original sound when they aren't (Ref. 19)

The surround-sound decoder uses the out-of-phase property to tell the difference between front-centre and rear-centre sounds, and it is also helped because the rear sounds are supplied with a frequency limit of only 7kHz. Thus the presence of high frequencies also tells the decoder they are meant to be at the front, and it routes them accordingly, even though they may consist of stereo music with many out-of-phase components. This technique also ensures that the effects of the outer ear are not brought into play, so listeners are not tempted to turn their heads. It is therefore an implicit assumption that the audience is seated facing forwards in a way which discourages turning around.

Setting-up the rear speaker(s) is quite a complicated business. To prevent tape hiss coming out continuously during the long periods of inactivity, the rear channel on "Dolby Stereo" recordings is further encoded Dolby B. (This is *in addition* to any noise reduction used for the distribution medium). Thus the decoder must be aligned so that the Dolby B threshold is correct (section 9.5). Furthermore, rear sounds should be delayed, so the front sounds arrive at the listener's seat first, and the Haas Precedence Effect (Ref. 20) does not destroy the front stereo image. Ideally, the delay should be about ten or twenty milliseconds at the listener's ears. In a large cinema this can be quite difficult to achieve, because sound takes about three milliseconds to travel one meter, and setting a delay which is appropriate for the entire audience may need several directional loudspeakers. But this is less difficult in a small room; the Dolby Surround circuit incorporates a delay control which, once set for a particular room, needs no further attention.

In the film world, it can also be assumed that the two channels (the two in the middle of the 4-2-4 system) are encoded Dolby A or Dolby SR on analogue release-prints (sections 9.5 and 9.13); I have ignored this additional complication here. "Dolby Stereo" is quite separate from Dolby Noise-Reduction.

Dolby Laboratories have since developed a digital audio process which can be added between the perforation-holes of analogue SR-coded prints (called "SR-D"), which may also include "Dolby Stereo" matrixed information.

I have mentioned all these details for a very good reason. I consider that it is *not* appropriate to make four-channel objective or service copies of a "Dolby Stereo" soundtrack, because the delay (and, for that matter, the gain) of the rear channel needs to be attuned to the listening-room in question. And because Dolby engineers were on hand during the mix to give advice about the inevitable consequences of a 4-2-4 matrix, it is better to store the sounds they way they intended them. All three copies - archive, objective, and service - should be two-channel versions. In my view, the "archive copy" should *not* incorporate decoded Dolby A or Dolby SR noise reduction where it exists, but the others should.

10.13 Developments of "Dolby Stereo" - (a) Dolby AC-3

A modified surround-sound system for domestic viewers has been introduced by Dolby Labs (Ref. 21); the audio is in a digital format. The system (called "AC-3") is a combination of several Dolby inventions, including their digital data-compression methods and the psychoacoustics of the "Dolby Stereo" system. With data-compression, Dolby originally claimed to be able to store their surround-sound information in a datastream of 320 kilobits per second. As with "Dolby Stereo" five full-range audio channels are encoded, plus another for low frequencies only (gunfire, aircraft, bombs, etc.). Dolby's data-compression technique employs masking psychoacoustic tricks, but these are augmented by another technique to reduce the bit-rate. When similar sounds are found in more than one channel, the coder economises on data by coding the information only once, registering the similarity. However, it became clear that 320 kilobits per second were not sufficient; Hollywood have been using 384 kilobits per second for their DVD movies, and Dolby themselves now recommend 448 kilobits per second. Fortunately, AC-3 decoders recognise the bitrate!

The geometrical layout seems to be similar to, and is claimed to be compatible with, cinema "Dolby Stereo." But instead of one channel of sound behind the listener, there are now two, to prevent the "point-source" effect if the listener should turn. The low-frequency sounds may be routed to one or more "sub-woofers", which (as they are often large) may be situated behind the viewer where he can't see them.

As described, this writer considers that "Dolby Stereo" and Dolby AC-3 are incompatible, because the cinema and domestic versions will need different source-mixes at the dubbing stage. In addition, the domestic version is in fact a "5½-1-5½" system (the trade press has changed this clumsy expression and calls it "5.1 channels.") Finally, it seems that the data-compression which results from steering single sounds would result in large "lumps" of data, whose sizes vary with the complexity of the scenes being simulated. It is not clear how this is implemented while maintaining synchronism with the pictures; but with digital video this is a relatively trivial problem. However, broadcasters have never liked it, because the licensing process means ongoing costs for television when broadcasters use it.

However, international politics have entered the arena, and DVD Video discs for the European market were originally mandated to use another bitrate compression system called MPEG-2. This was an "open" system (so anyone could use it; royalties are only payable on hardware); but delivery of the necessary processing software and hardware was about a year behind AC-3, which therefore dominated the first year of DVD. MPEG-2 is used for some European DVDs, and many DVD players are capable of decoding it; but AC-3 is now "the compulsory option" for DVD Video. Yet another system (called "DTS" – Digital Theater System) appeared in 1993.

In the case of DVD, all this was feasible because the physical disc carries the necessary software to decode the bitstream it bears. The implications for archivists, I leave for the reader! As I go to press, the MPEG Forum has now developed a non-backwards compatible version of surround sound called "AAC" (Advanced Audio Coding). It is not the purpose of this manual to predict the future; so I shall stop at this point.

10.14 Developments of "Dolby Stereo" - (b) Dolby headphone

This system, launched in the spring of 2001, is a way of modifying conventional analogue Dolby Surround so the effect may be heard on headphones. There are no transducers picking up movements of a listener's head, so the circuit (which is an analogue one requiring special processing chips) can feed many headphone listeners at once. It emulates one of three sizes of room-space around the listener; where an entire audience is watching the same film (on an aircraft, for example) only one processor is necessary (Ref. 22). From the point of view of a sound archive, you would certainly need to provide "bypass circuitry" for ordinary stereo or mono listening.

10.15 Developments of "Dolby Stereo" - (c) Pro Logic 2

This, also introduced in the spring of 2001, is an upgrade to the normal analogue Pro-Logic decoder for stereo sounds (section 10.12). It *must* have a "centre front" loudspeaker, and routes dialogue psychoacoustically into that loudspeaker (also Ref. 22). The mono "rear signal" now becomes one with a wider frequency response, and is "spread" to avoid the point effect. Apparently it also offers a "music mode" to generate surround-sound for listening in a car.

10.16 Discrete 4-channel systems - The JVC CD-4 system

The following four sections deal with "4-4-4" systems, beginning with two which were distributed on vinyl LP discs and meet the definition of "quadraphonic" given at the start of section 10.7.

"CD-4" was a discrete quadraphonic system invented by the Japanese Victor Company in 1970 (Ref. 23). More than 500 LPs were issued in the next five years, the chief protagonists being JVC themselves, and the RCA and WEA companies of the USA. A few highly specialist issues appeared (for example, some steam-train noises, which conventional stereo couldn't handle adequately).

The channels comprising the left-front and left-back sounds were matrixed using conventional "sum-and-difference" techniques. The sum-signal was cut onto the left-

hand groove wall with the usual equalisation in the usual manner, so the special CD-4 pickup could also be used for playing conventional stereo discs. The difference-signal was encoded using JVC's "ANRS" noise reduction system (section 9.7), and modulated an ultrasonic carrier at 30kHz. (This was actually frequency-modulated below 800Hz and phase-modulated above 800Hz, the total passband being 20-45kHz). The right-front and right-back channels were cut on the right-hand groove wall in a similar manner. Reproduction used a pickup with a response to ultrasonic frequencies, fed directly to a CD-4 decoding box. (Some quadraphonic amplifiers were also available with a decoder built-in). After the decoder had been matched to the pickup and stylus (an alignment disc being necessary), this simply reversed the foregoing processes, and provided four channels of audio.

I haven't written the above paragraph "just for the record". The process was so complex that I cannot believe any archivist would build a CD-4 decoder from scratch. But I need to make the point that no standard digital recording system has a frequency-range wide enough to permit a "warts-and-all archive copy" as defined in section 1.5. For this and other reasons, it seems that present technology may force us to break our rule, and convert the sounds back into four channels for storage in digital form.

Here are some of the "other reasons." The ultrasonic-carrier principle strains analogue disc technology to the limit; it is much less likely that vinyl bearing such grooves will have a long shelf-life. Pickup cartridges with a response to 45kHz and good crosstalk figures are essential, and it is much less likely that these will be available in future decades. Even if they were, supplies of "Shibata" diamonds (section 4.11) will also be scarce. Finally, it is known that one playing with a normal stereo pickup is enough to wipe the information from the ultrasonic carrier, and the longer we leave it the more likely this is to happen.

I mentioned that CD-4 alignment discs were necessary for setting up the gear; the only one I actually know is Panasonic SPR111 (which is a 7-inch "45"), but there must have been others.

I haven't mentioned a certain important principle for some time; but as always, I recommend that the archivist should start by locating the four-track master-tapes. If this isn't possible, then I recommend that CD-4 discs be decoded and transferred to 4-channel digital recordings as a matter of some priority. These would then combine the functions of "archive copy," "objective copy," and "service copy."

10.17 The "UD-4" system

A quadraphonic disc system. The basic idea was a hierarchy of matrixes invented by Duane H. Cooper in 1971, and was developed by Nippon Columbia for LP discs from 1973 onwards. The proponents called it a "4-4-4" system, but this is flattering; there was inherently a certain amount of crosstalk, although always uniform round the square. Despite some optimistic announcements, very few LPs were actually made. The only example I know is the Hi-Fi News "Quadrafile" album of May 1976, comparing the four different systems QS, SQ, CD-4 and UD-4 on four sides.

When announced in mid-1974, UD-4 was described as combining features of several quadraphonic disc systems in an attempt to make a "universal" one, but later revelations showed that it was "just as unique" as its predecessors. The groove contained a true omnidirectional sound channel (rather like Ambisonics, in the next section), and a second difference-channel. This was also omnidirectional except for a frequency-

independent phase-lag that, in comparison the first channel, was made equal to the panorama bearing angle for each of the images to be portrayed. It is not clear how this was supposed to represent two or more sources at once; but on single-point sources the crosstalk was something like that of the regular matrix (-3dB for neighbouring loudspeakers and -infinite for diagonal loudspeakers). It would *not* have resulted in a stereo-compatible audio signal, but no-one seems to have noticed this point. After de-matrixing, the theoretical "emission polar diagram" for a point source of sound reproduced at the centre of a square array of identical loudspeakers was shaped like a cardioid. The ultrasonic channels improved this in much the same way that two cardioid microphones can be combined to give a "second-order cardioid" response. It was possible to reduce the bandwidth of this second-order signal without much effect on the subjective result, and the UD-4 carrier did not swing beyond 40kHz. (Ref. 24)

Archivists may like to know, however, that Denon did make a "universal decoder/demodulator" (their Model UDA-100). This would decode QS, SQ, CD-4 and UD-4 without any of the "enhancing" circuits, so this is a machine to be sought by sound archivists. (Ref. 25)

10.18 "Ambisonics"

"Ambisonics" is a hierarchical system for capturing and reproducing sounds in various directions with respect to the listener. Although this isn't how it's actually achieved in a practical session, the "core" of the system may be imagined as an omnidirectional microphone, picking up sounds from all around. (Conventional terminology calls this the "W" channel). Associated with this are up to three additional signals representing the directionality in the three dimensions. (The same convention calls these X (front-back), Y (left-right), and Z (up-down), and if the complete sphere is recorded this way on a four-channel medium, it is called "B-format.") All these signals are assumed to have been picked up at the same point in space. Time differences do not enter into Ambisonics, so the system is inappropriate for headphone listening. Given perfectly-engineered microphones, the system picks up sounds with the correct polar diagrams, with no crosstalk or dead zones, unlike previous systems.

Ambisonics also comprises means for manipulating audio captured using this basic principle. Such signals may be stored or transmitted on two, three, or four audio channels. Two channels might permit the effects of "stereo," three channels might permit the effects of "quadraphony," and four channels might permit the reproduction of height information as well. All these uses of the word "might" are because the results would depend on how the channels were manipulated. Furthermore, loudspeakers are not be confined to any idealised geometrical layouts; the reproducing system could reconfigure the channels to *any* loudspeaker layout without compromising their information-content. So ambisonics allows infinite flexibility of directional reproduction. (Ref. 26)

So far, all the pre-recorded software has been two-channel for stereo listeners, usually encoded with a matrix called UHJ. I have not been able to find any account of how the UHJ process works, but I gather this too is hierarchical, enabling two dimensions to be matrixed onto two channels and three onto three. A review of the matrixing process considered alone on a B-format signal (encode and decode) is given in Ref. 27. This is the only example of a matrix being reviewed "before and after," and an implication of this review was that the UHJ matrix was not supposed to compromise the directionality of any

of the original four-channel images (unlike other matrixes). It did, though! Once again, the first step must be to locate four-track Ambisonic master-tapes.

10.19 Other discrete four-channel media

I write this section with some diffidence. It seems to me that the existence of a four-channel version of any recording should not go unremarked in a programme of conservation-copying. The extra spatial information should be preserved as well as the best power-bandwidth product. The media in this section are almost certain to have inferior power-bandwidths, but will carry spatial information uncorrupted by matrixing processes. Considerable discographical research will be needed to discover what was going on.

In the USA there were a number of specialist manufacturers of pre-recorded four-track quarter-inch tapes, giving a reasonable-quality 4-4-4 quadraphonic distribution medium for material recorded by various commercial companies. Unfortunately, these were often on acetate-based tape, needing priority copying to preserve the sound.

There was also a "superior" version of the eight-track endless tape cartridge for cars. Instead of having four sets of stereo tracks, this had two sets of quadraphonic tracks. A few were made and sold in Britain, the most important being some of EMI's best-selling pop albums from the years 1975-1980. Also there was some quadraphonic material which was never available in any other form, notably from the British Decca companies. Unfortunately, the power-bandwidth product of an eighth-width quarter-inch tape running at 9.5cm/sec was very poor.

Whilst these media might represent the intended position of musicians more accurately than a matrixed disc of the same session, we must also remember that the sound-balancers may have been monitoring through a matrix encoder/decoder system. The sales of matrixed discs were usually insufficient to pay for themselves, so it is very unlikely that another mix would be done specially for discrete tapes. I frankly do not know whether this avenue would be worth researching, nor can I say which medium would best match the producers' intentions.

The official Philips "Red Book" standard for compact digital discs also has an option for 4-channel sound. Nobody has yet taken this up as it implies CD players with four digital-to-analogue converters, although both hardware and software was promised for the second half of 1982. (Ref. 28). But I mention this because it offers the only hope of resolving the difficulty. If commercial discrete 4-channel CD players become available, an archive should pursue them as a matter of principle.

10.20 More than five-channel systems

Greater numbers of channels seem to have been confined to film media, although I suppose it's always possible they could crop up anywhere. For example, "Cinerama" had seven discrete channels. It was claimed that it used seven separate microphones at the shooting stage, and the seven tracks were reproduced on nine loudspeakers (five behind the screen, one "on each side of the orchestra", one at the back of the auditorium, and one at the back of the balcony). (Ref. 29)

So far I have been unable to find whether similar spatial resolution was actually captured by other systems. "Todd-AO" used six tracks, and the Russian "Kinepanorama"

system used nine; but it is at least possible that these were made up from one- or three-channel master recordings with suitable panning, simply as the easiest way of ensuring the sounds came from the right places when reproduced.

When I was working with wildlife filming, the "height" dimension proved essential. I greatly regret that only one system (Ambisonics) has ever offered this, although I developed something using six bi-directional microphones around a cube of absorbent material; this also preserved some "time information".

10.21 Multitrack master-tapes

I shall describe the effects of multitrack audio recording upon musicians in section 13.5; but here I shall concentrate on the issues affecting an archivist. In all the previous sections, I have concentrated upon reproducing the sound from analogue media "faithfully" (that is, reproducing the sound the way the original engineers *intended*); but in nearly every case, a multitrack tape implies subjective treatment at a mixing console. Although the analogue tape standards I described in Chapter 6 imply only a problem with "hardware", there are no international standards in the digital domain, so both "software" and "hardware" is needed for playing most digital multitrack media - let alone the "artistic" elements of the mix-down. Even *digital* multitrack equipment implied an analogue mixing console until roughly the year 2000.

Unless a satisfactory hardware-based multitrack digital medium is used, there is no way in which a digitised version of an analogue multitrack tape can be preserved in a satisfactory manner (section 3.7). This has a serious side-effect for students of musical history. What happens to performances which were never "mixed-down" at the time?

The only two solutions I can see are as follows. One is to play the unreleased performances into a sound-mixer set with the controls at "flat", so students will at least get some idea of what is on the multitrack master. The other is to employ a mixing operator familiar with the subject matter, and who can "imitate the style of" the original mixes. This would enable the *character* of the music to be emulated, rather than a "warts-and-all" reduction. And - preserve the originals!

REFERENCES

1: Michael Gerzon, "Surround-sound psychoacoustics," (*article*): *Wireless World* Vol. 80 No. 1468 (December 1974), pp. 483-486.
2: Christina Morgan, "Thorn-EMI audio move," London: *Broadcast*, 22nd October 1993, p. 16.
3: Barry Fox, "Sound waves in sync for better stereo," London: *New Scientist*, 23rd October 1993, p. 20.
4: IBS News (The Journal of the Institute of Broadcast Sound), No. 22 (November 1985), pp. 16-19.
5: Barry Fox, "Business", *Studio Sound* Vol. 32 No. 11 (November 1990), p. 32.
6: European Patent Application 357 402.
7: Barry Fox, "TV Audiences in 3-D," *Studio Sound* Vol. 33 No. 7 (July 1991), pp. 58-59.
8: Michael H. Gray, "The Arturo Toscanini Society" (*article*), Journal of the Association of Recorded Sound Collections, Vol. V No. 1 (1973), page 29.
9: Barry Fox, "Pairing up for stereo," *Studio Sound* Vol. 28 No. 3 (March 1986), p. 110.

10: L. F. Rider, "Magnetic Reproduction in the Cinema" (*article*), Sound Recording and Reproduction (the Journal of the British Sound Recording Association), Vol. 5 No. 4 (February 1957), pp. 102-5.
11: *Wireless World*, December 1972 p. 597.
12: Geoffrey Shorter: "Four-channel Stereo - An introduction to matrixing," *Wireless World* January 1972 pp. 2-5 and February 1972 pp. 54-57.
13: P. Scheiber, "Four channels and compatibility," *J. Audio Eng. Soc*. Vol. 19 (1971) pp. 267-279. (Presented at the AES Convention 12th October 1970; also reprinted in "Quadrophony" Anthology, AES 1975, pp. 79-91.
14: "Quadraphony and home video steal the Berlin Show," *Wireless World*, vol. 77 (1971) pp. 486-8.
15: R. Itoh, "Proposed universal encoding standard for compatible four-channel matrixing," *JAES* April 1972. (First presented at the 41st A.E.S Convention, 7th October 1971). Also reprinted in "Quadraphony" Anthology, AES 1975 pp. 125-131.
16: B. B. Bauer, D. W. Gravereaux and A. J. Gust, "A Compatible Stereo-Quadraphonic (SQ) Record System," *J. Audio Eng. Soc*., vol. 19 (1971) pp. 638-646. Also reprinted in "Quadraphonics" Anthology, AES 1975, pp. 145-153.
17: Geoffrey Shorter, "Surround-sound Circuits - Build your own matrix circuits using i.cs," *Wireless World*, March 1973, pp. 114-5.
18: (Haas precedence effect)
19: Simon Croft, "Europe Is Surrounded - Using Dolby Surround: how does it affect production?" (*article*). TVBEurope, October 1994, pp. 38-9.
20: P. S. Gaskell, B.A., and P. A. Ratcliff, B.Sc., Ph.D., "Quadraphony: developments in Matrix H decoding" (*monograph*). BBC Research Department Report RD 1977/2 (February 1977).
21: Barry Fox, "Video viewers to surround themselves with sound," *New Scientist* No. 1819 (2nd May 1992), page 20.
22: Barry Fox, "Dolby's 'phones and PL2", London: One to One (*magazine*), January 2001 page ?
23: T. Inoue, N. Takahashi, and I. Owaki: "A Discrete Four-Channel Disc and Its Reproducing System," *J.A.E.S*. vol. 19 (July/August 1971). Originally presented at 39th AES Convention, 13th October 1970; and reprinted in "Quadraphonics" Anthology, AES 1975, pp. 162-169.
24: Duane H. Cooper and Toshihiko Takagi, "The UD-4 System," *Hi-Fi News & Record Review*, Vol. 20 No. 3 (March 1975), pp. 79-81.
25: Gordon J. King, "Equipment Review: Denon UDA-100 Decoder/Demodulator," *Hi-Fi News and Record Review*, Vol. 20 No. 8 (August 1975), pp. 117-8.
26: Richard Elen, "Ambisonics - Questions and answers," *Studio Sound*, Vol. 24 No. 10 (October 1982), pp. 62-64.
27: Peter Carbines, "Review of Calrec UHJ Encoder," *Studio Sound*, Vol. 24 No. 9 (September 1982), p. 88.
28: Mike Bennett (Sony Broadcast Ltd.), "Letter to the Editor," *Studio Sound* Vol. 23 No. 9 (September 1981), p. 51.
29: D.W.A. (= almost certainly initials of Donald W. Aldous), "Cinerama" (*article*), *Sound Recording and Reproduction* (the Journal of the British Sound Recording Association), Vol. 4 No. 1 (December 1952) page 24.

11 Dynamics

11.1 Introduction

We now enter upon an aspect of sound restoration which is almost entirely subjective, and therefore controversial - the restoration of original dynamic range.

People do not seem to realise that the range of sound volumes in a musical or dramatic performance was usually "squashed" when the performance was recorded. It is a great tribute to the skills of professional recording engineers that so many people have been fooled. But you should learn how engineers altered things in the past, so you know how to do your best if you are ever called upon "to restore the original sound."

The restoration of full volume range is never appropriate for an "archive copy," there have to be special circumstances for an "objective copy," and it is rarely essential for a "service copy." But I maintain that a present-day operator with experience in controlling certain types of music when *recording* it, might be best-placed to *undo* the effect on similar music (and this applies to other subject matter, of course). It will nearly always be a subjective issue, and I mention it only because if the approach is to reproduce the original sound as accurately as possible, a present-day operator could be better-equipped to do the job than any future operator, as he knows the craftsmanship of masking the side-effects.

The same principles can be used when we consider that past engineers always worked with a definite end-medium in mind, whether it was acoustic disc reproduction, A.M. radio, or compact disc. Not only would they have kept the signal within suitable limits for such applications, but they would also have taken into account the "scale distortion" I mentioned in section 2.12. So, on the assumption that future listeners will not be working with the same medium, we may have to allow for this on the service copy at least.

There are three levels of difficulty. At the first level, manual adjustments of volume were made while the recording took place. At the second level, electrical apparatus was used to provide automatic compression of the dynamic range, and the role of the engineer was to ensure that the apparatus didn't affect the music adversely. At the third level, automatic apparatus was used which was "intelligent" enough not to need an engineer at all; it could select its parameters depending on the subject matter, and could generally reduce volume ranges more than the first two levels.

Which brings us to a definite reason why I must write this chapter. Suppose your archive has made an off-air recording, and the broadcaster uses your archive so he may repeat something. Then you *must* reverse the "second and third levels" above, or the effects will be doubled. This neither represents the wishes of the original producer, nor makes a broadcast *you* can be proud of (let alone the listeners). Therefore you *have to* reverse the effects!

Recovering the sounds is rather like using a reciprocal noise reduction system (Chapter 8) whose characteristics are unknown. In fact, it is sometimes possible to confuse recordings made using reciprocal noise reduction with recordings made using other types of dynamic restriction. Your first duty is to ensure the recording was *not* made with reciprocal noise-reduction! A useful side-effect of restoring the original dynamic range is that background-noise of the original analogue media may be reduced at the same time.

Hardly any of the necessary apparatus for restoring dynamic range has yet been developed, although relatively minor modifications of existing machinery could allow most of the possibilities. Restoring the original range means moving the background-noise of any intervening medium up and down as well. This may prove a distracting influence which fights your perception of the original sound. It may be best to work with a copy of the recording which has already been through a sophisticated noise reduction process (sections 0 to 4.21). Since many of these processes are only appropriate for a service-copy, you might as well "be hung for a sheep as for a lamb" and get better dynamics as well!

I hope this essay will inspire others to do some development work. I seem to have been working on my own in this field so far. No-one else seems to have recognised there *is* a problem; yet there has never been much secrecy about what was going on. The techniques were taught to hundreds of BBC staff every year, for example. Nevertheless, I will start with a statement of the basic difficulty.

11.2 The reasons for dynamic compression

Since the beginning of time, it has been necessary to control the dynamic range of sounds before they came to the recording medium. Such controlling may be "absolute," that is the sensitivity of the machine may be fixed with no adjustment during the recording, or "relative," in which the overall sensitivity *plus* the relative sensitivity from moment to moment, are variable. Such control has always been necessary, because practical analogue recording media are unable to cover the complete dynamic range perceived as sound by the human ear. (Note that, once again, I am having to imply human hearing; readers concerned with other applications must make the appropriate mental adjustments). To oversimplify somewhat, there may be 120 decibels between the faintest detectable sound and the onset of pain, whilst practical recording media with this dynamic range are still only at the experimental stage. Furthermore, we cannot yet reproduce such a range, even if we could record it. And when it becomes technically feasible, health and safety laws will stop us again!

Therefore recording engineers have always practised implicit or explicit dynamic controlling, which usually means some distortion of the quantity of the original sound on playback. The next few sections will therefore comprise brief histories of different aspects of the topic, followed by my current ideas about how the effects of such controlling might be undone, or rather minimised.

Undoing such compression means you need lots of amplification. Please remember what I said above - I cannot be held responsible if you damage your loudspeakers or inflict hearing loss upon yourself. I am quite serious. There is a very real danger here, which will be exacerbated because many of the clues needed to control such sound lie at the *lower* end of the dynamic range. So you will have to turn up your loudspeakers to hear these clues. I also advise very quiet listening-conditions; these will make listening easier without blowing your head off (or annoying others). And, although you may find it fights what you are trying to control, I strongly advise you to install a deliberate overloading device in your monitoring system (such as an amplifier with insufficient power), preferably arranged so it comes into effect at the maximum safe volume. That way you will be able to separate the extra anomalies, because they will not coincide with distortion of the monitoring system itself.

11.3 Acoustic recording

There were only two ways of controlling signal volume in the days of acoustic-recording. The absolute sensitivity of the machine could be adjusted by changing the diaphragm at the end of the horn. Professionals used easy-to-change diaphragms for that very reason. The power-bandwidth principle came into effect here, although it was not recognised at the time. A thick diaphragm, being relatively stiff, had a high fundamental resonant frequency resulting in good treble response, but had low sensitivity to most notes within the musical range. A thin diaphragm, being relatively compliant, gave inferior high-frequency response, but greater sensitivity to lower notes.

This shall be considered in greater detail in the next chapter, but in the meantime I shall discuss the "relative" adjustments. The only techniques for altering sound volumes during a "take" were to move the artists, or to use artists trained in the art of moderating their performances according to the characteristics of the machine. The balance between different parts was predetermined by a lengthy series of trial-recordings. There was no way in which the absolute or relative sensitivities of the apparatus could be adjusted once the recording had started (Ref. 1). Because of these factors, it is the author's view that we have no business to vary the dynamics as they appear on acoustically-recorded discs or cylinders. They were determined by the artists themselves, or by the person with the job of "record producer," and for us to get involved would subvert the original intentions of the performance.

11.4 Manually-controlled electrical recording

There is no doubt at all that from the earliest days of electrical recording, volume adjustment was possible by means of the "pot" (short for "potentiometer"), better known as the "volume-control." This introduced a variable electrical resistance into the circuit somewhere, and the principle is still used today. I shall use the term "pot" for the artefact concerned, whether it actually comprises an electrical resistance with a wiper contact, or a sophisticated integrated circuit controlled through a computer, or a combination of any number of such devices. The essential point is that, in my terminology anyway, a "pot" will imply that there is someone with a hand on a control somewhere, so the signal volume can be controlled before it gets to the recording machine.

This paragraph outlines the history of the hardware. Pots started off as wire-wound variable resistors with fairly fine resolution from one turn of the coil to another (usually much less than a decibel), but suffering from a tendency to make noises as the volume was changed. Professionals from about 1930 to about 1970 used "stud faders," in which the wiper moved over a set of contacts made from precious metals to resist corrosion. Precise fixed resistances were wired to these studs. They were designed to give a predetermined degree of attenuation from one contact to the next, and the physical layout of the studs ensured consistency, so the same results could be guaranteed if it was necessary to change to a different fader, or use duplicate equipment, or repeat a session. There was always a trade-off between a large number of contacts and their resolution. Eventually most organisations settled on "2dB stops," and engineers might refer to changing the sound by "four stops," meaning eight decibels. Values between these 2dB steps could not be obtained except by sheer luck if a wiper bridged two contacts. Line-up tones could be set up on a meter this way, but the process couldn't be trusted in real recording situations. With some types of sound (low frequencies with a heavy sinewave

content), audible "sidebands" occurred as the wiper moved from one stud to another. It was perfectly obvious to those concerned (who usually reacted by cleaning the studs vigorously, an inappropriate action since the clicks were an inherent side-effect of the system); but it is worth noting this, since it shows when volume-controlling is *unlikely* to have taken place. The effect could be concealed by making the change occur on a wideband sound, during a pause, or at a "new sound" such as a bar of music with new instrumentation or dynamics.

Meanwhile, semi-professionals and amateurs used "carbon track pots," comprising a shaped arc of carbon compound with a wiper running along it. This gave infinite resolution, but was prone to inconsistency (especially noticeable in the early days of stereo), and electrical noise if the track wasn't perfectly clean. From about 1970 professionals preferred "conductive plastic" faders, which could be made consistent and free from sidebands. Sometimes the problem of dirty contacts was ameliorated by using the pot to control a "VCA" (voltage controlled amplifier).

The purpose of the pot (or a combination of pots and/or switches and/or integrated circuits) was threefold. First, it set the "absolute level." This ensured the voltages were in the right ballpark for the subsequent apparatus. Next, the operator manipulated the pot so that the loudest sounds did not have any undesirable side-effects, such as distortion, or repeating disc-grooves, or aural "shocks" which might alienate the audience. Thirdly, the operator manipulated the pot to ensure that the quietest sounds did not get lost in the background noise of the recording medium or the expected listening environment, or were too weak to have the subjective impact that the artists desired. It will be seen that there was a large subjective element to this. It wasn't simply a matter of keeping electrical signals within preordained limits.

Nevertheless, the electrical limits were important, and pots were almost always operated in conjunction with a meter of some sort, sometimes several at once. The design of such meters is not really a subject for this manual, but you should note that there were several philosophies, and that certain authorities (such as the B.B.C) had standardised procedures for operating pots in conjunction with such meters. As archivists, it may be appropriate to treat B.B.C recordings using B.B.C metering principles if the intended effect of the broadcasting authority is to be reproduced. (Ref. 2). Unfortunately, I have not been able to find an equivalent document for any other such organisation, although aural evidence suggests that something similar must have existed at one or two record companies.

The important point is as follows. Nearly all sound operators used both ears and eyes together to set the signal before it was recorded. Ears were needed to listen out for technical and artistic side-effects and strike a compromise between them, and eyes to read a meter to inform the operator how much latitude there was *before* side-effects would become audible.

Such side-effects might not be apparent to the engineer immediately associated with the performance. Meters also helped users of the sound further "downstream." They ensured that broadcasting transmitters at remote locations didn't blow up for example, or that disc recording machines didn't suffer repeating grooves, or that tape recordings could be spliced together without jumps in absolute volume.

The present-day operator will have no difficulty distinguishing between two quite different sound-controlling techniques. One is the "unrehearsed" method, in which the engineer was unaware what was going to happen next, and reacted *after* the sound had changed. For example, a sudden loud signal might create an overload, and the engineer would react *after* the event to haul the volume down.

The other is the "rehearsed" method. The action here depended upon the signal content. If the idea was to preserve the *contrast* between the quiet and the loud sounds, for example to preserve the effect of a conductor's sforzando without overloading the recording, the engineer might nudge the volume down carefully in advance, perhaps over a period of some tens of seconds. But if the idea was simply to keep the sound at a consistent subjective volume, for example two speakers in a discussion, he would wait until the last possible moment before moving the pot; nevertheless, it would be *before* the new speaker started talking. Double-glazed windows were provided between studio and control-room so he could *see* when a new speaker opened his mouth or a new instrument took a solo; but an intelligent engineer could often do the same job without the benefit of sight by his personal involvement with the subject matter. So a present-day operator ought to be able to reverse this action when necessary.

The point is that the present-day restoration operator should be able to distinguish between the "unrehearsed" and "rehearsed" methods of control. In the former case, we will have violent and unintended changes of sound level which cannot be called "restoring the original sound," and they should be compensated as far as possible on the service copy. In the latter case, the contemporary operator was working to produce a desired subjective effect. Whether we should attempt to undo this depends, in my view, on whether the performance was specifically done for the microphone or not. If it took place in a studio at the recording company's expense, then we should preserve the dynamic range as the company evidently intended it; but if the microphone was eavesdropping at a public concert, then it could be argued that the dynamic range as perceived by the concert audience should be reinstated.

11.5 Procedures for reversing manual control

At present we have no way of setting the "absolute level" of a sound recording unless the recording happens to come with a standard sound calibration. (And see section 2.12). So we are confined to reversing the "relative" changes only.

There are two main indications operators can use for guidance. One is if there is a constant-volume component to the sound, such as the hum of air-conditioning equipment. By listening to how this changes with time, it is possible to use this clue to identify the points at which the original engineer moved his pot, and by how much. The background need not be strictly constant for this to be possible. Passing traffic, for example, comprises individual vehicles coming closer to the microphone and receding again. However, there is always a certain rate-of-change to this, which does not vary so long as the traffic is moving more-or-less constantly (which a present-day listener can identify by ear). Thus, when an anomalous rate-of-change is perceived, it is a sign that the original engineer may have moved a pot or done something that needs our attention.

The other case is where the wanted sound itself has a change of *quality* accompanying a change of *quantity*. The human voice is a good example of this; voices raised in anger are different in quality from voices which are not raised, and even if the original engineer has done a perfect job of holding the recorded signal at a constant electrical value, the change in quality gives it away. The same applies to orchestral music. This writer finds, in particular, that both strings and brass playing fortissimo have more harmonics than when they play mezzo-forte. It is less obvious than the spoken-word case; but by skipping a disc pickup across a record and listening to quality and quantity together, anomalies often show up. The original engineer was able to conceal these

anomalies by operating his pot slowly or very quickly, as we saw earlier. By comparing different passages in rapid succession, his basic philosophy will become clear, and it can then be reversed.

Obviously, the present-day operator must have a very good idea of the actual sound of a contemporary performance. He might have to be a frequent concert-goer, and a practicing sound-controlling engineer, and be familiar with hearing uncompressed performances on his loudspeakers, if his work is to have any value. But provided these requirements are met, the present-day operator could be better-equipped to tackle this challenge than anyone else.

The principal trap is that dynamic controlling may have been performed by the artist(s) as well as the engineer(s). Maybe the orchestra played mezzo-forte rather than fortissimo to make life easier for the engineers. This frequently happened in recording studios, and in British broadcasting studios prior to the advent of "studio managers" in the early 1940s. A meter cannot be used for undoing this effect. We can never say "this mezzo-forte was 15 decibels below peak fortissimo on this modern recording of the work, so we will make this old performance match it on the meter." If the operator does not take the harmonic structure of the sound into account, the resulting transfer may be a gross distortion of the original. I believe that if an orchestra restricted its own dynamic range for any reason, the service-copy must reflect that.

There are a few other clues which can sometimes be exploited to identify dynamic compression. It is well-known that, with the advent of the potentiometer, engineers frequently controlled the volume *downwards* as a 78rpm disc side proceeded from the outer edge to the inner, in order to circumvent "inner-side distortion" (section 4.15). This almost always happened whenever a disc was copied, since potentially-distorting passages could be rehearsed. Most reasonable recording-venues have reverberation which decays evenly with time; indeed, the standard definition of "reverberation time" is how long it takes the reverberation to decay by sixty decibels. A listener may be able to spot when a change of volume has occurred, because there will be a corresponding apparent change in reverberation-time for a short period. Indeed, if he knows the sound of the original venue, he will be able to recognise this effect even when it occurs throughout the *whole* of a particular recording, perhaps because an automatic compressor was used. Which brings us to the next topic.

11.6 Automatic volume controlling

In the early 1930s automatic devices became available for restricting the dynamic range of sounds before they were recorded or broadcast. Today these devices are called "limiters" and "compressors." Different writers have given different meanings to these two words, so I shall follow the dominant fashion, and define them my own way!

I shall use the word "limiter" to mean a device which existed as an overload-protector, and a "compressor" as a device which had some other effect (usually artistic). I will also use the word "expander" for an automatic device which can increase dynamic range. The archivist can use expanders to undo some of the effects of compressors and limiters, but a full range of facilities does not yet exist on any one commercially-available device.

Limiters were invented first, and provide most of the work for us today. They started in the film industry, to help control spoken dialogue, and protect the fragile light-valves then used for optical film recording (section 13.14). With the relatively limited

signal-to-noise ratio of the medium and suitable manipulation of other parameters, it was found that four to eight decibels could be taken off the peaks without the effect being apparent to anyone unfamiliar with the original. Thus the limiter quickly became another tool for creating convincing soundtracks from unconvincing original sounds. For films at least, an "unlimiter" should not generally be used today; but, if (say) a recording of film music were to be reissued on compact disc, it could be argued that an unlimiter is justified because the sound is being copied to a different medium.

I have found no evidence for limiters in pure sound work before 1936, when the R.C.A Victor record company used one to help with Toscanini's uncompromising dynamics on commercial orchestral recordings. By 1939 they were commonplace in the US recording industry, and the B.B.C adopted them from 1942 to increase the effective power of their A.M. transmitters under wartime conditions. As with the film industry, four to eight decibels of gain-reduction was found useful without too many side-effects being apparent to listeners unfamiliar with the originals. But it must be remembered this judgement was made in the presence of background noises higher than we would consider normal today. If we succeed in reducing the background noise of surviving recordings, it may require us to do something with the dynamics as well.

An engineer generally listened to the effects of a limiter for the following reason. When two sounds occured at once, one might affect the other, with a result called "pumping." We shall come to some examples, and how the matter was resolved, later; but the best way of overcoming this difficulty in early days was to have two limiters, one for each sound, and mix the sounds after they had passed through the limiters.

By 1934 Hollywood had separate limiters on the dialogue, the music, and the effects. The final mix would have another limiter, but essentially this would function as an overload-protector; the component tracks would have been controlled individually. In sound-only media, again the RCA Victor company seems to have been at the forefront, with several limiters being available on each of the microphones as early as 1943.

The B.B.C also incorporated limiters in most of their new disc recording equipment from this time onwards. Here we have a real ethical dilemma, because there was always one limiter in circuit for each pair of recording machines (called "a channel", designed to permit continuous recording). It was generally pre-set to take 4-8 dB off the peaks, and was not an "operational" control. Surviving nitrate discs may have a much better background-noise than that perceived by contemporary A.M listeners, which can be an argument in favour of undoing the effect. On the other hand, A.M listeners might have heard it after passing through a second limiter at the transmitter, this being an argument in favour of *doubling* the amount of expansion today! You can see the ethical difficulties; clearly, whether we should undo these effects will depend sharply upon the subject matter, whether it was also transmitted on F.M. or only on wartime A.M., the purpose for which the disc was made (e.g. was it an "insert" into a programme, or made for the Transcription service or given to the artist as part of his fee), and other considerations.

The European record industry was strangely slow to adopt limiters, and even slower to adopt several at once. The earliest example I know is EMI Columbia's recording of Act 3 of *Die Walküre* at the 1951 Bayreuth Festival, where the device was used specifically to compensate for the fact that the singers were moving about on stage. During the same season Columbia covered *Die Meistersinger*. In this opera, the commercial 78s show a limiter was used during the disc-cutting process, but not on the LP (or the recent C.D). So by this time, limiters were already being used in two different ways.

With the advent of rock music from 1954, limiters became almost mandatory in popular music. Rock music was meant to be LOUD. Limiters not only enabled 100% loudness, they also enabled the unpredictable antics of vocalists to be balanced against the comparatively steady backing. Whenever mixing is taking place *after* one or more limiters, the limiter is an essential part of the balance, and we should not try to undo it.

When F.M radio started in Europe in the mid-1950s, limiters functioned simply as overload-protectors. F.M was supposed to be a hi-fi medium. Unnecessary distortions were abhorred, and limiters at F.M transmitters were hardly ever triggered until the advent of commercial local radio in Britain in the mid-sixties.

Limiters were sometimes incorporated in domestic and portable recording equipment from about 1960, this was usually to act as an automatic way of setting the recording-level, thereby making recording easier for non-professionals. Such limiters had widely varying characteristics depending upon the imagined customer-base. A cassette-recorder for the home might react relatively slowly to overloads, but would eventually settle down to an average setting which would still allow loud peaks to sound loud. However, dictation-machines for offices would even out consecutive syllables of speech when delivered on-mic, while not amplifying office background noise, or being upset by handling-noise or mike-popping. We should therefore study contemporary practices if we aim to undo the effects; fortunately, this type of equipment was rarely used for archivally-significant work.

However, limiters became a real menace in domestic and semi-pro video equipment. The distractions of the picture-capture process meant automatic sound-control became almost essential. This wasn't a problem in the film industry, because professional Nagra machinery became the industry standard. It incorporated a very "benign" design of limiter, which tended to be used carefully by professionals with rehearsals, and its results were always subject to further control in the film dubbing theatre. But amateur and professional video engineers tended to ignore the audio aspects of their kit, often using it without rehearsals in the presence of noisy backgrounds to record syllabic speech with fast volume changes. In addition, the noises made by the camera and its crew, and the difficulties of the sound-man (if any) hearing what was happening under these conditions, means most video sound is a travesty of the original.

A final twist, which happened even with films, is that "radio mikes" became available from about 1960 to mitigate the problems of microphone cables. Usually, such mikes were used "in shot" (e.g. by an interviewer or a pop vocalist); but sometimes an experienced operator can tell a radio-mike has been used even when it can't be seen, because it would be impossible to cover a scene any other way. (For example, long tracking shots across soft ground, or actuality battle-scenes). Radio mikes always incorporate limiters so they conform to international radio regulations when the sound gets too loud.

In the 1970s the battle for increased audiences meant more sophisticated signal-processors were used to increase the subjective loudness of radio stations. The manufacturers had proprietary trade-secrets, but generally the devices used more complex attack-time and decay-time circuits. Also the frequency range might be divided up into several separate compressed "slices," so that a loud note in one slice did not affect the sound in another, and the overall volume remained high without perceptible side-effects. It is my unhappy duty to tell you that I do not yet know a way of reversing the effects of these devices. In principle, it is possible to "reverse-engineer" the effects of most of these devices. We might obtain the required parameters by comparison with (say) untreated commercial records which have been broadcast by the same station, or by acquiring an

example of the actual device and neutralising its effects with op-amps, or by something which once actually happened to me. After I had done a live radio broadcast, the studio engineer proudly gave me a tape of what I'd said (knowing I was doing my own off-air recording at home), with the words "This *hasn't* been compressed". But for the remainder of this chapter, I shall confine myself to the second level of difficulty I mentioned above, automatic devices with no "intelligence."

11.7 Principles of limiters and compressors

It is necessary to understand the principle of operation of a limiter before we can reverse-engineer it. I shall start with a "normal limiter." This classic design is also known as the "feed-back" limiter, and was the only practical design before the mid-1960s. After that, devices with different architectures became possible, thanks to the development of solid-state VCAs (voltage-controlled amplifiers) with stable and predictable characteristics which didn't need the feed-back layout.

The circuitry had the architectural shape depicted in Fig. 10.1.[4] Uncontrolled electrical sounds arrived from the left, and volume-setting would usually be achieved by a pot P. This might or might not be an operational control, depending on the reasons for the presence of the limiter. There would now follow a piece of circuitry which gave a variable degree of amplification or attenuation. It has been shown here with the letters "VCA"; although this is a modern term, the principle is much older. The degree of amplification could be varied by changing a "sidechain" voltage; here, this voltage is depicted as coming up from below.

What is important is how this control voltage varied. This is what we must emulate to restore the original dynamics today. In a "feed-back" limiter, a sample of the *output* voltage was taken. In Fig. 10.1, it is being amplified in a stage shown symbolically as "A"; however, most limiters actually combined the amplification and VCA functions in the same piece of hardware. These have been shown separately, because the amount of amplification affected the compression-ratio of the limiter. The signal was then rectified (converted to a direct current) and applied to an analogue .AND. gate. Again, this is a modern term, although the principle is much older. An .AND. gate only passes signal when voltages are present at *two* inputs. In this case, a fixed voltage is present at the other input of the gate, here shown coming from a battery. (In practice, most limiters had a special power supply for this purpose, which was made variable in some designs). Only when the voltage at the first input exceeded that of the second did any voltage appear at the output of the gate. The output then travelled back to the VCA as the control voltage.

Thus, the limiter behaved as a normal amplifier until the signal reached a particular level of intensity. After this, the control voltage reached the VCA, and the amplification was reduced.

The degree of reduction (or the "compression-ratio"), was dependent upon the amplification at A. Since the VCA-stage often incorporated its own amplification, this is not a definable quantity in many practical limiters. With lots of amplification, no matter how much louder the signal may have become at P, the output would increase barely at all. With less gain, the limiter would turn into a "compressor" - a device that reduced peaks in proportion to their size, rather than bringing them all down to the same fixed level.

[4] Figure 10.1 not unavailable

Fig. 10.2[5] depicts some steady-state characteristics of devices with "feed-back" architecture. Curve (A) is the "ideal limiter" characteristic, never actually achieved in practice. Curve (B) shows what was often accepted as a satisfactory limiter with "ten-to-one" compression, and curve (C) shows the actual performance of a 1945 Presto recording limiter for discs. Curve (D) shows how reducing the amplification at A might give a reduced compression-ratio - in this case, "two-to-one." This would be for subjective reasons, rather than for preventing overloads.

The control voltage was made to change with time by a number of components, the three most important being shown symbolically in Fig. 10.1[6] by R1, R2, and C1. These represent two electrical resistances and a capacitor in the side-chain. Usually, if there was an overload, the aim was to get the signal volume down quickly before any side-effects could happen (distortion, repeating grooves, etc). So the limiter had to act quickly. We shall consider this factor in section 11.9. After the peak had passed, it was necessary to restore the amplification so quieter sounds would not be lost. C1 and R2 served the function of restoring the status quo; any charge built up on capacitor C1 would leak away through R2, and the amplification of the VCA would slowly revert to normal. This function will be considered in section 11.10.

Three other parameters are relevant. One is to identify the action of the .AND. gate. If the limiter was being used as an overload protector, this setting would obviously be left alone for the duration of a recording or broadcast, or else the level of protection would vary. Practical limiters might have a control called "Threshold" or "Set Breakaway" to adjust this; but the gain of the subsequent recording equipment also had the same effect, and cannot be quantified today. We shall need to identify the whereabouts of the threshold by watching the compressed signal on a peak-reading meter.

The second is the compression-ratio. Curve 10.2B shows ten-to-one compression. At first, we might try using one-to-ten expansion to restore the original range. Unfortunately, this is usually impossible to achieve for two reasons. First, our threshold-emulator has to be set extremely accurately, and this must remain true at all frequencies. Half a decibel of error will result in five decibels of incorrect expansion, and an audible disc click may be amplified into something like a rifle-shot. Secondly, the actual expansion-ratio was never consistent, as we can see from curve 10.2C; this too can result in wild deviations with small causes. An expansion-ratio of about three-to-one is about the best we can reverse-engineer with present technology. But all is not lost; we shall see how to circumvent these problems in section 11.11.

The third parameter is "side-chain pre-emphasis." There might be an audio equaliser incorporated in the amplifier A. The most common reason for this was to increase the amount of treble going to the rectifier. In an F.M transmitter, for example, the wanted sound is transmitted with high-frequency pre-emphasis, as we saw in section 7.3. If the high frequencies are allowed at their full volume they will overload (or break international regulations). The equivalent operation could, in theory, be done by putting the pre-emphasis circuit *before* the VCA; but then it is difficult to judge the subjective effect.

Oddly enough, side-chain pre-emphasis seems rarely to have been used on mono disc records, despite the presence of pre-emphasis. But when motional feedback cutterheads became available in 1949, there was a sudden demand for the facility. Motional-feedback cutterheads have their resonant frequencies in the middle of the audio range, so kilowatts of power might sometimes be needed to force the mechanism to trace

[5] Figure not available
[6] Figure not available

an undistorted high-frequency wave. And when you burnt out a motional-feedback cutterhead with its special coils, the damage was very expensive. There was *always* a sidechain-preemphasis limiter on stereo cutterheads, although every effort was made not to trigger it. Sidechain-preemphasis was normal on optical film recordings, for much the same reason.

Now for some words on the subjective use of the facility. Normally, vowel sounds yield higher voltages than consonants (particularly sibilants), so using a "flat" limiter would compress the vowels and leave the sibilants proportionally stronger. Side-chain pre-emphasis diminished this effect. Optical film sound was very prone to high-frequency blasting, and the matter was not helped by cinema soundtracks being reproduced more loudly than natural. As far as I know, all limiters for protecting optical light-valves had side-chain preemphasis, with different settings for 35mm and 16mm film. This circuitry comprised of a resonant peak rather than a normal 6dB/octave equaliser. This had the advantage that short transients, such as gunshots, did not cause the limiter to respond immediately, and thus the transient quality was preserved. Many early microphones and cutterheads had noticeable resonances in the range 5kHz to 10kHz. A naturally-pronounced sibilant could be turned into an irritating whistle by such equipment. Sidechain preemphasis could diminish the irritation without affecting the other consonants, so its effect was much appreciated by speakers with false teeth!

11.8 Identifying limited recordings

The first task of the present-day operator is to establish *whether* a limiter was used, so we know *when* to reverse-engineer it. The best clue is to reproduce the recording with the correct equalisation, and watch it on an analogue peak-programme meter (PPM). This is the standard BBC instrument I mentioned above in section 11.4. It has a mechanical pointer with a relatively short attack time and a long recovery time, and it is possible to see exactly how the peaks on the recording behave to infinite resolution (unlike modern "digital" or "bar-graph" displays). If the peaks resolutely hit the same point on the dial and never go beyond it, a limiter has almost certainly been used. Un-limited speech and music usually has at least one "rogue peak" a couple of decibels louder than any other.

Another sign is if the sounds *between* the peaks are heard to "pump", as we defined in section 11.6. This is very conspicuous when there is a constant element in the wanted sound - for example, air-conditioning noise or background traffic. Pumping is also noticeable when there is a number of peaky things sounding at once. A choir with eight treble voices is a good example. Treble voices are especially prone to "peakyness" with their vibratos, and several such voices often modulate each other in an easily-recognised manner. Applause can have a similar result, especially if there are one or two dominant individuals. (And again I must remind you not to confuse this effect with that of a reciprocal noise reduction system).

A third clue is whether the subject matter was unrehearsed, so the operator was forced to use a limiter with a faster-than-human reaction-time to prevent the troubles we mentioned before. The recordings of the British Parliament, with rowdy and unpredictable interjections during the debates, are prime candidates; we can tell a limiter has been used even though no side-effects are audible, because it would be quite impossible to cope any other way!

You can sometimes detect a limiter by its effect on starting-transients; this gives a characteristic "sound" which you will have to learn, but I shall be enlarging upon this point later.

The best way for an original recording engineer to avoid pumping was to combine several separate microphones which were limited (or compressed) individually. Thus compression was an integral part of the mix, and should not be undone. The rigorous peaking of the PPM does not tell you whether this has happened; you must use your ears and your experience.

Perhaps the best way of learning the effects of a limiter is to use one - or rather, misuse one - yourself; but do it on material which hasn't been controlled before! Most broadcasts and many LPs have already been limited by operators skilled in minimising the side-effects, and such material will not give typical results. As I keep saying, any skilled *maker* of such recordings is supremely well-equipped to *unmake* them.

11.9 Attack times

Having proven that we *do* need to neutralise a limiter, we should first think about setting our equipment to emulate the action of R1. This is how we do it.

If R1 was a very low resistance, there would be nothing to prevent the limiter operating quickly, which was the desired aim. But practical circuits always had some resistance, so R1 always had a finite value, even when the circuit designer didn't actually provide a physical resistance R1 in the wiring. Thus all "feed-back" limiters had a finite "attack-time." It was usually in the range from ten milliseconds to 0.1 milliseconds. Anything longer than ten milliseconds might give audible harmonic distortion or catastrophic failures of powerful amplifying equipment, and two or three milliseconds was the most which could be tolerated for preventing repeating grooves in the constant-amplitude sections of disc-recording characteristics. Even-shorter attack-times were needed to prevent the clash of ribbons in optical light-valves.

In the 1940s and 1950s, it was noticed that starting-transients (particularly the starts of piano notes) sounded wrong. At first, this was attributed to "overshoot." It was believed that the start of the note was passing through the limiter without being attenuated, so there was a short burst of overloading in the following apparatus until the limiter attacked. So limiters were made with even shorter attack-times, but matters got no better.

Eventually it was realised that the sudden change of volume was generating sidebands, like stud faders, and the solution was deliberately to *slow* the limiter. The BBC fixed on ten milliseconds for its transmitters as being the optimum for balancing the side-effects of distortion and sidebands; but this was too long to prevent groove-jumping on discs, and the recording industry generally used shorter attack-times. I shall not go any further into the philosophy of all this. In conclusion, if you im to "undo" a limiter, you may need to emulate the attack-time today in certain circumstances. And the best way of doing this is - once again - to emulate R1 using an experienced ear.

Until now, I have not said anything about the box marked "VCA." In the days of valves, this comprised a "variable-mu" valve, a device whose amplification was directly controlled by a grid in the electron path. Unfortunately, a change-of-voltage coming from R1 would be amplified and emitted as a "thump." To neutralise this, professional circuits used a pair of variable-mu valves in push-pull. The valves had to be specially selected and the voltages balanced so the thump was perfectly cancelled. Unfortunately, this did not

always work. In domestic applications, the only solution was to increase the attack-time or cut the bass at the output, so the thump was effectively infrasonic. The Fletcher-Munson effect often masks quieter thumps; but they can be shown up by watching the waveform on an oscilloscope, when the zero-axis appears to jump up and down.

Although I have done a lot of work in "unlimiting" recordings, I have not yet concocted a circuit to generate an equal-but-opposite thump. When we get this right, it will provide another clue to setting the attack-time and the degree of expansion, because we shan't get consistent results otherwise; and when we *do* get consistent results, it will be a powerful proof we are doing the job objectively. But this is in the future.

The cases of very long attack-times are fortunately much rarer, and easy to detect. Sometimes the machine took as much as a second to react, so that it sounds like a clumsy human-being behaving in the manner we saw earlier. However, it will do this consistently. It will not have the common-sense to realise another loud peak may follow the first, and this will tell you an automatic device is responsible!

At this point, I must mention that in 1965 came the "delay-line limiter", a device which had a new architecture differing from the classical layout. This stored the sound for a short time in a delay-line, taking the side-chain signal from *before* the delay-line, so the VCA could reduce the gain *before* the peak got to it. This architecture is known as a "feed-forward" limiter, and can have other advantages (such as infinite compression-ratio). But for the moment, please remember that from 1965 onwards we may find recordings with *negative* "attack-times."

Fortunately, the setting of the attack-time is not often a critical matter today. Most sounds are of relatively long duration, longer than about ten milliseconds. So most limited peaks will be followed by a "plateau" of sound of more-or-less constant volume, during which time most expanders will respond satisfactorily. Only extremely short loud sounds (such as hand-claps and the starts of piano-notes) will cause trouble.

11.10 Decay-times

To reverse the "decay-time," we must emulate the setting of the R2/C1 combination. This is usually more important than emulating R1. The decay-time defined how long it took for normal conditions to be re-established after a peak. Engineers had several ways of defining it, which I won't bother you with. But it had to be considerably longer than about a tenth of a second, or appreciable changes would happen during each cycle of a loud sound, causing harmonic distortion. The maximum encountered is usually around ten seconds or so; I cannot be more precise without going into how to define this characteristic. Fortunately, we don't need to bother with this complication. It never seems to be documented, so we have to set it by ear anyway.

In practice, the decay-time of the limiter or compressor was usually made adjustable to suit the subject matter being recorded. A fifth of a second corresponded to syllabic speech, so the limiter could compress one loud syllable without affecting others. At the other extreme, ten seconds was slow enough not to affect the reverberation behind music.

A further complication occurred when a limiter was installed as an overload-protection device which might have to cope with *all* types of subject matter. A broadcast transmitter is an obvious example; transmitter engineers certainly couldn't re-set the switch every time speech changed to music. Furthermore, even at the long setting, there was always the risk of an isolated peak. One loud drum-beat would then be followed by

many seconds of music at reduced power. Although this could be solved by two limiters in series set to different recovery-times, the solution adopted was a minor modification. All BBC transmitters and the Type D disc-cutters had this facility (although not the Presto disc-cutting equipment); and when the idea was fitted to commercial limiters from about 1955 onwards, it was described on the switch as "AUTO" or some such word. I call this a "dual recovery-time."

To set the recovery-time of our expander we must usually listen to the aural clues I mentioned in section 11.5. We shall probably discover why there were so many definitions of "recovery time"! The VCA frequently did not have a linear response to the control voltage. Although the voltage will have decayed exponentially at R2, the VCA-section may have resulted in a different recovery-pattern. We will probably find a satisfactory setting for reversing small peaks, but be unable to hold any constant background-noise steady between louder or wider-separated peaks. At present, I do not know an expander which enables us to correct for this, although it is a relatively trivial matter to build one. Something with an adjustable non-linear transfer-characteristic is needed in the side-chain.

A further development was the "triple recovery time." (Ref. 3). This was developed especially for the all-purpose transmitter protection I mentioned earlier. The third time constant was made dependent upon the *output signal-level*; when this fell below a certain level for a certain amount of time, the machine assumed a new speaker was about to start talking, and reset itself quickly. In the writer's experience this worked very well. For the record, the new circuit was installed at the first ten BBC Local Radio stations in 1969.

11.11 The compression-ratio and how to kludge It

The next thing we shall discover is that we cannot know, much less reverse-engineer, the exact compression-ratio of the particular limiter. It is rather like balancing a razor-blade on its edge. Even if we knew the threshold, the attack-time, the decay-time, *and* the actual characteristic, any slight mismatch will have a catastrophic effect - usually the gain of the expander will increase dramatically and blow up your loudspeaker. In any case, you can see from curve 10.2(C) that there may be ambiguities. If we were trying to simulate this particular limiter and we had an output 4dB above the threshold, we would not know if this corresponded to an input of +5dB or +20dB. Usually we have to live with the fact that a limiter destroys the very information needed to expand the signal again. In this respect, it differs from a reciprocal noise reduction system, where the compressed sound is deliberately designed to be expandable.

How then do we undo a limiter at all? Here is where we make use of the other clues I listed earlier - any steady background-noise, the harmonics of musical instruments, etc., - routing the signal through an expander circuit.

This does nothing until a peak signal comes along. The circuit detects such a peak because the threshold at the .AND. gate is set by a control R4. However, the circuit then expands the signal, not according to any pre-set expansion-ratio, but according to the setting of another pot R5. This is where the most important part of the subjective judgement comes in.

The operator will need a great deal of rehearsal to set this pot, and he will need to keep his hand on it all the time as the work progresses. He will also need a marked-up script (for speech), or a score (for music). By listening to the aural clues, he will

continuously manipulate this pot according to the loudness of the original sounds, and the circuit will take over responsibility for pulling the sound up faster and more accurately than the operator can do it manually (where 0.1 to 10 milliseconds was the norm). It will also do the chore of restoring the original gain at the correct recovery-time after the passage of each peak, while the operator sets R4 ready for the next peak according to his rehearsed script.

Another useful facility is to insert a filter at Q. Almost as soon as the sound film was invented, cinema-goers noticed that a "flat frequency response" was not successful. There was a great deal of discussion whether this should be the responsibility of the studio or of the cinema, but the psychoacoustic factors were not realised at first. A definitive answer did not arise until 1939, when "dialog equalization" was researched, to compensate for cinema audiences hearing dialogue from the loudspeakers at volumes much louder than natural (Ref. 4). The result was to cut the bass somewhat. For television and video that has been through a limiter but not a "dialog equalizer", the original sound may be re-created more faithfully by preceding the expander with a treble-lift circuit and following it with a treble-cut circuit, with reciprocal responses so the overall frequency response is maintained. The researches of Ref. 4 have proved very helpful in developing such circuits. The result is that the signal is expanded more as voices are raised in pitch. Since louder voices are higher in pitch, background sounds can be made to sound consistent for longer periods of time.

These ideas form, frankly, two kludges to get around anomalous and inconsistent compression-ratios; but they often work. The only time they cannot work is when the recording is compressed so deeply that some of the *background* noises reach peak volume. If the *background* reaches the setting of R4, the circuit will overemphasise it to the *peak* setting R5. This is particularly troublesome when the recovery-time R2 is short. (When it's long, it's possible to switch the expander out-of-circuit manually before the background reaches peak volume).

One potential solution is to introduce a filter into the side-chain at X which discriminates between the wanted and unwanted sounds. If the unwanted sounds are low-frequency traffic-noises for example, it is sometimes possible to insert a high-pass filter here which allows speech vowel sounds through, but cuts the traffic. This does not eliminate the traffic frequencies from the corrected recording - the original sounds are preserved - but it prevents them triggering the expander anomalously.

Another idea (which I have *not* tried) is to insert a highly selective filter at X. If there is a reasonably constant background, such as the hum of air-conditioning, this can be selectively picked out, and used to control both the time and the amplitude of the expansion process. But, as I say, I haven't tried this and I am sceptical regarding its success. Automatic selection is difficult when the hum is in the *background*. I suspect the human ear is better able to pick out such features when there are loud foreground sounds. However, expanders with such side-chain filters are available commercially.

Thus, to cope with the problem of backgrounds and foregrounds being of similar volume, I prefer a digital audio editor in which it is possible to pre-label passages where expansion should take place (and by how much), so it will ignore passages where the background must *not* be expanded. Research into such a program is taking place as I write.

REFERENCES

1: Peter Ford, in his article "History of Sound Recording" (*Recorded Sound* Vol. 1 No. 7 (Summer 1962), p. 228), refers to the surviving acoustic lathe at EMI He says: "A cord and coiled-spring tensioning device . . . provided a means of tuning the resonances of the diaphragm assembly to fine limits." This was a misunderstanding. About twenty years after that article was published, the cord perished, and the assembly fell apart. Thanks to the co-operation of Mrs. Ruth Edge of the EMI museum, I was able to examine the parts, and it was possible to see that the cord-and-spring assembly was designed to permit the quick exchange of cutting styli. It could not be used to tune an individual diaphragm, much less alter its properties during a performance.

2: British Broadcasting Corporation: C. P. Ops. Instruction No. 1 - "Control and Modulation Range Instructions." This version came into effect on 1st January 1957, and was reprinted, with additional information, in "Programme Operations Handbook (Sound Broadcasting)", December 1956, pp. 165-170.

3: British Broadcasting Corporation: D. E. L. Shorter, W. I. Manson, and D. W. Stebbings, Research Department Report No. EL-5, "The dynamic characteristics of limiters for sound programme circuits." (1967). This was reprinted in a slightly shortened form as BBC Engineering Monograph No. 70 (October 1967).

4: D. P. Loye and K. F. Morgan (Electrical Research Products Inc.), "Sound Picture Recording and Reproducing Characteristics" (*paper*), Journal of the Society of Motion Picture Engineers, June 1939, page 631.

12 Acoustic recordings

12.1 Introduction

This chapter introduces sound recordings made without the help of any electronic amplification. Such recordings were dominant before 1925, and the technology was still being used in remote locations without electricity until at least the late 1940s.

To oversimplify somewhat, the performers projected their sounds into a conical horn, which concentrated the sound waves onto a small area, and vibrated a tool which cut a groove in wax. In a few cases we know that professional "experts" utilised a few decibels of *pneumatic* amplification, and I shall take a brief look at this in section 12.25. But the vast majority used simpler apparatus which is shown diagrammatically in section 12.3. Virtually everything I am talking about will be the result of an "unamplified" performance.

We are currently at the leading edge of technology in recovering "the original sound" from acoustic recordings, and significant amounts of experimental work are taking place as I write. In the mid-1990s this author wrote a series of five articles on the subject, which (since it forms the first reference for nearly everything which follows) I shall cite now to get it out of the way! (Peter Copeland: "Acoustic Recordings" (*series of articles*), Sheffield: "The Historic Record" (*magazine*), nos. 32 (July 1994) to 36 (July 1995)).

I am very grateful to the Editor of "Historic Record", Jack Wrigley, for permission to make use of substantial sections of those articles; and the resulting correspondence continues to shed much light. In that series, I attempted to outline how one might achieve the correct equalisation of acoustic recordings, using much the same principles as I described for electrical records in Chapter 6. Since then, I have had the privilege of meeting one or two other operators who accept that faithful reproduction of acoustic recordings might be possible, and new ideas and projects have been started (and sometimes abandoned).

With the current state of the art, it is my duty to report that it is impossible *exactly* to prescribe the correct treatment for an acoustic recording. But I can foresee that intelligent computer software will eventually replace the human hearing process with its implied subjective judgements. So this chapter outlines the current state of the art in anticipation of further developments; and, to save space, it does not describe the numerous experiments to prove or disprove the theories. (Many such descriptions were included in the above Reference, and will continue to be made available to serious enquirers).

12.2 Ethical matters

Before 1925, recording engineers were called "experts". Discographical studies show that in a surprising number of cases (perhaps one in four), we know the identity of the expert concerned. It has frequently been said that experts were individuals with techniques all their own, and therefore we cannot expect to reverse-engineer their products today. Frankly, I think this is rubbish - contemporary image-building, in fact. We can and we must, otherwise we will never learn the full story.

But I agree it might *not* be ethical to compensate for acoustic recording machinery. I am trespassing upon the subject of the next chapter, but contemporary experts were well aware of the deficiencies of their equipment, and managed their recording-sessions to give the optimum musical result. When we reverse-engineer the recordings, more often than not we get a horribly *un*-musical result. But I also hope this chapter will demonstrate how and why defects occurred, so we may at least respond intelligently and sympathetically to those recording artists whose legacy seems disappointing today.

There is also a great deal of evidence that professional artists modified their performances before the recording-horn. To deal with one consideration: I remember I was almost overwhelmed by the amount of sound which came back at me when I first spoke into such a horn. A vocalist might notice his A-flats were reflected back strongly, for example; so he would deliver his A-flats with reduced intensity (and possibly modified tone-colour).

The experts were forced to do everything they could to get a loud enough signal above the surface-noise. Yet complete symphony orchestras, concertos, and choirs were attempted - or, to use a better word, simulated. For there is no doubt that the less-than-perfect reproducing equipment of the time enabled artists and experts to get away with faking sounds. This therefore raises the moral dilemma of *whether* we should try to reproduce the original sounds.

The earliest acoustic recordings were made with machinery so insensitive that there was hardly ever any intended sound audible on playback. This remains generally true today, even with modern restoration technology. Pioneer experts concentrated on making the performers operate as loudly as possible; so early recording locations were chosen more for their advantages in not annoying the neighbours than any other reason. Secondary considerations might be warmth (to help keep the wax soft enough to cut), and height (to allow a reasonable fall for the weight-powered motor). Thus early recording studios tended to be at the top of buildings. They were often lit by skylights. They happened to provide lighting similar to the studios used for photography, painting, and sculpture, which made performers slightly more at ease; and it was easy to close the skylights during a "take", and then reopen them to let the perspiration dry out.

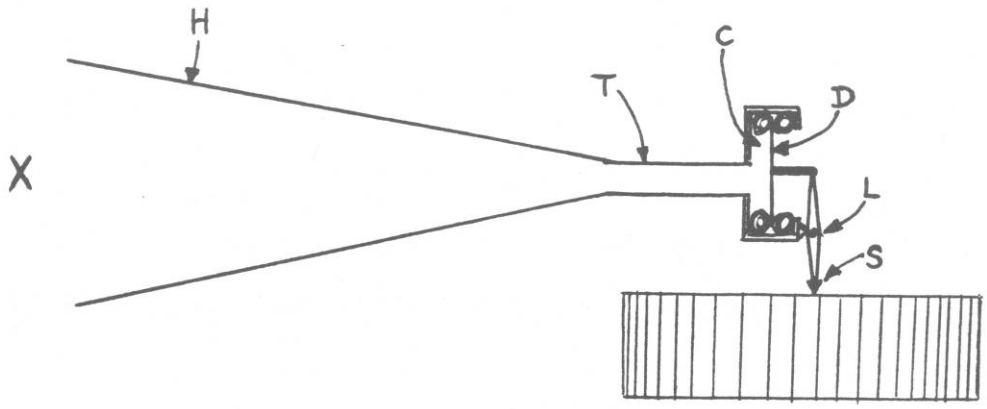
From a very early date, the advantages of sound reflecting surfaces were appreciated. A reflective surface behind the performer could help drive more sound into the horn to operate the recalcitrant mechanism. (The technique of the "hard screen" is still used in professional recording studios, of course). As acoustic recording technology became better at dealing with groups of artists, it was found that the average stuffed Victorian or Edwardian room was a hindrance, so the nature of the surrounding furnishings radically changed. Hard bare walls and floors and ceilings were introduced. I have no doubt that experts tried different sizes of rooms; but anything greater than about five meters square and three meters high was found to be actually disadvantageous. A bigger room might have given musicians and music-hall artists a more familiar environment, but the acoustics would not have been picked up by the horn. A much smaller room with reflective walls would be like a jazz band in a telephone box - extremely loud, but quite impossible for satisfactory performances. A slightly larger room would suffer, because the bodies in it would mop up most of the sound reflections; whereas the room I have just described hit an acceptable balance between reflected sound being powerful enough to augment the signal, while the bodies in the room were insufficient to mop up this advantage. Someone once described it as "making music in a sardine tin," and when we make an objective copy of almost any acoustic recording today, I am struck by how accurate this description is.

I can think of only one acoustic recording where the surrounding environment played a significant part - the recording of the Lord Mayor of London, Sir Charles Wakefield, in 1916. (Ref. 2) In this recording it is just possible to hear a long reverberation-time surrounding the speech, which must have been overwhelming on location. Apparently the Gramophone Company's experts thought this was so abnormal that they took the precaution of insisting that "Recorded in Mansion House" was printed on the label. As far as I can ascertain, this was the first case where the location of a commercial recording was publicly given, although we now know there were many previous records made away from the usual studio premises.

There is yet another reason for writing this chapter, which we haven't come across before. A large number of acoustically recorded discs and cylinders survive which document the dialects and music of extinct tribes and peoples. Nearly all experienced sound operators are familiar with the special sound of acoustic recordings, and can mentally compensate for the distortions caused by the recording horn. But less experienced listeners, who are used to the sound of microphones, sometimes find it difficult to adapt. To avoid misleading these people, it is advisable to give them processed service copies with the worst distortions reduced, even though this may not be to the usual degree of precision. It is particularly important for the non-professional performers on such records, who did not modify their performances to suit the machinery. The relationships between different vowel sounds for example, and between vowels and consonants, may be seriously misunderstood.

12.3 Overall view of acoustic recording hardware

Most acoustic recordings were made by apparatus shown diagrammatically below. Sounds would be emitted by a performer at X; this represents the position of the principal vocalist or soloist. There might be several performers at different positions around him, behind him, and beneath him; but the recording equipment was so insensitive that anyone not in the optimum position X would be relegated to a distant balance in the recording.



The sounds would be collected by means of a horn H, usually in the shape known to students of Euclid as "the frustrum of a right cone with a semi-included angle between

eight and eleven degrees." (I shall call this "conical" for short !). The purpose was to concentrate sound from a large area at the mouth of the horn upon a much smaller area. Quite often several horns were provided for several principal performers. Sometimes a parallel-sided tube T followed, and sometimes a cavity C serving the function of an acoustic impedance-matching transformer, and sometimes both.

The sound waves then passed to a unit known as the "soundbox" or "recorder box", where they vibrated a flat circular diaphragm D mounted between compliant surrounds. Both sides of the diaphragm were exposed to the atmosphere, and the difference in pressure between the two sides caused it to move. (A physical screen between the artists and the expert would modify this situation, of course). The diaphragm's centre, where the maximum movement normally occurred, was usually connected to a lever system L which carried the vibrations to a cutting stylus S. These vibrations modulated a spiral groove cut into a solid disc or cylinder made of wax. Some sort of lever system seems always to have been used, even on hill-and-dale recordings. I mention this because in theory it would be possible to couple a stylus directly to the diaphragm in hill-and-dale work; but a linkage permitted the stylus to be changed more easily, and allowed for variations in the level of the wax.

There were many variations on the above scheme, but it accounts for more than half all acoustic recordings, and the variations do not greatly affect the results. I shall consider the components of the "standard" system first, leaving the variations until section 12.25 when they will make more sense to you.

As we consider each component in turn, we shall learn where our understanding has reached today. Virtually all this understanding will be in the frequency response of the recording machine. I don't want to befuddle you with mathematics, so I shall describe the effects of each component in words, and give references to scientific papers for further study.

There is some evidence that harmonic distortion occurred, but most of this "evidence" seems to come from writers who did not distinguish between the "blasting" which occurred during recording and that due to reproduction. With the exception of hill-and-dale recordings which were not transferred from a "master" (and which therefore did not have their even-harmonic distortion cancelled), I have rarely heard significant distortion recorded into the groove (see Section 4.15). Although such distortions always happened to a limited extent, they nearly always seem to be outweighed by reproduction distortions, and I shall not consider them further.

This is the point to remind you that if the signal-to-noise ratio of the reproduced recording is good, it is more likely to result in hi-fi sound recovery. But it is particularly important in this chapter, because the insensitivity of the recording-machine resulted in large amounts of noise at both ends of the frequency range. We must therefore apply most of the noise reduction principles discussed in Chapter 4. Also the frequency response had "steep slopes." When we compensate for the frequency response, we also "colour" the background noise; and this may even corrupt the process of determining the right equalisation in the first place, as we shall see.

12.4 Performance modifications

Drawn from various written sources, here are details of some of the compromises made during commercial recording-sessions. The piano always caused difficulties, because the sounds came from a sounding-board with a large area. In early days it was always an

upright, which was also easier to get close to the horn. The back was taken off so there would be nothing between the sounding-board and the horn, and the pianist was instructed to play "double forte" throughout. Sometimes the piano's hammers were replaced with something harder than felt, because of the diaphragm's insensitivity to transient sounds. There was little improvement in the diaphragm's sensitivity after about 1906. But better records (with less surface noise) and better gramophones (with greater acoustic efficiency) meant that the apparent sensitivity increased, and by 1910 classical pianists could be recorded playing relatively unmodified grand pianos.

Orchestras were also difficult. Stringed instruments always gave problems, because they had less power than the wind section, and their operators needed more "elbow-room." The Stroh Violin was commonly employed; this was a violin fitted with a soundbox and horn so the upper partials of violin tone could be directed at the recording machine. Yet even with this artificial clarity, contemporary string players were encouraged to exaggerate the portamento and vibrato features of their instruments to help distinguish them from the wind. As bass notes were recorded very weakly, the bass line was often given to the tuba, which had more partials; these could be recognised on playback as constituting a "bass line."

So brass or wind bands would be preferred for vocal accompaniments. Stage-management was needed during a "take" to give soloists uninterrupted access. Vocalists had to bob down out of the way during instrumental breaks. Because of the bass-cut difficulty, those unfamiliar with recording technique had to be man-handled by the recording director to bring them close on low notes and push them away on high notes.

A healthy amount of reverberation will help musicians, because it helps them to stabilise their sense of pitch. It is a vital part of every trained singer's singing-lessons that they should be able to judge the pitch of their singing by a combination of three techniques: voice-production method, which ensures the pitch is as perfect as possible before the note starts; hearing the sounds through their own heads to provide a "tight feedback loop"; and hearing sounds reverberating back from around them to provide longer-term stability and match what other musicians are doing. The sense of pitch is not directly related to frequency. It varies with volume in a complex way, and the trained singer must balance these three pieces of evidence. An untrained one will rely almost completely upon the third piece; hence the phenomenon of "singing in the bath." (I suppose I had better explain this remark, now that bathrooms usually have thick carpets and double-glazing. In the first half of the twentieth century, bathrooms were bare places, with reverberation-times longer than any other room in a private house; so that is where aspiring singers would try their voices).

The most notorious example of someone upset by wrong acoustics was Amelita Galli-Curci, who found the relatively small studio and the loud accompaniments of the Victor Company's acoustic recording studio very difficult. She was frequently sharp in pitch. History doesn't record the "takes" which never reached the processing-bath, nor the trial-recordings which must have been made so she could adapt to the alien conditions; but the surviving evidence sometimes shows twenty or thirty processed takes of each song. Galli-Curci was a celebrity for whom this was worthwhile; but this is one explanation why books like Brian Rust's "Discography of Historic Records on Cylinders and 78s" and "The Complete Entertainment Discography" are filled with important artists who made test-recordings and nothing more.

Although I have little written evidence to support my next remarks, I strongly believe that when the medium of sound recording began to dominate over sales of sheet-music, techniques in instrumentation and song-writing changed. The popular song

became a "three-minute sound-byte" until liberated by the 45rpm "disco single" in 1977, for example. But, more fundamentally, the acoustic recording process was unable to handle both low frequency and high frequency sounds, so the "bass and drums section" of bands before the 1930s would have been impossible to record effectively.. So the rhythm was also forced to be in the middle of the range, and it had to be a simple rhythm. Thus it was given to the banjo and/or piano, and fox-trots and waltzes dominated. I have even seen it written that, in acoustic days, "the bass drum was banned from the recording studio"; but I can think of one or two exceptions. Modern reproduction methods are beginning to reveal such broken rules.

12.5 Procedures for reverse-engineering the effects

If our aim is simply to "reproduce the original sound," it may reasonably be asked why we do not simply measure the performance of surviving acoustic recording equipment to quantify it. There are several reasons.

Firstly, contemporary amateurs frequently modified phonographs (or accidentally damaged them) in such a way that the performance would be very significantly affected. In the professional field, only one complete machine survives at EMI Hayes (Peter Ford, "History of Sound Recording," Recorded Sound No. 22, p. 228). I owe a deep debt of gratitude to the EMI archivist, Mrs. Ruth Edge, for allowing me to make a very close visual inspection of the equipment, and my dependence will be clear in later sections where its description overshadows everything else.

Secondly, the performance of all such equipment depended critically upon the properties of perishable materials such as string and rubber, which have altered with time.

Thirdly, some experts used their own personal soundboxes and recording horns which were their own trade-secrets. Indeed, we do not always know which *expert* made which recordings, let alone which *equipment*.

Fourthly, the *way* the equipment was used - the placing of artists relative to the horn - also makes rigorous measurement largely irrelevant.

Finally, some measurements have actually been done, but the results are so complex that full *understanding* cannot be gained.

Because we generally know little about the equipment, most successful work has proceeded on the basis of getting what we can from the record itself. Yet we must not ignore the few exceptions to this idea. These fall into two classes.

One is where we know that *the same* equipment was used for two or more recordings (or sessions). This enables us to take the common features of both and use them as evidence of the machine which made them. For example, Brock-Nannestad used a particular horn resonance to distinguish between Patti singing a song transposed into another key, and the speed of the turntable being altered.

The other is extremely important. The Gramophone Company of Great Britain used standardised equipment for all its recording experts from 1921. Works of music longer than the four-minute limit were featured about this time, and standardised equipment permitted any one side of a multi-side set to be retaken at a later date with a quality which matched preceding and succeeding sides. George Brock-Nannestad has explained how the US Victor Company (who had a majority shareholding in Gramophone) were unhappy with their matrixes imported from Europe, so they sent one of their recording experts (Raymond Sooy) to reorganise the recording operation (Ref. 1). From that date,

the Gramophone Company's "Artists' Sheets" documented the equipment actually used for each take in coded form.

12.6 Documentation of HMV acoustic recording equipment

The "Artists' Sheets" show both published and unpublished recordings made by the Gramophone Company between 15th March 1921 and approximately February 1931, but they exclude recordings for the Zonophone label. They may be consulted at the British Library Sound Archive. British and Red-label International recordings are on microfilms 360 to 362; Vienna and points east, 385-6; Italy, 386-7; Spain, 387-8; France 388-9; other countries, 390-1.

The acoustic sessions are distinguished by *not* having a stamped triangle after the matrix number (which would mean a Western Electric recording). Three columns at the right-hand end of the sheets document the equipment used. The order of the columns and their contents are not absolutely consistent, but due to various redundancies (as we would say nowadays), there are no ambiguities.

We lack a "Rosetta Stone" to enable us to decipher the coded information. Ideally my deciphering should be regarded as a preliminary suggestion, which needs testing by other workers on other sound recordings; but I have checked the results for self-consistency in a number of ways.

The first column generally contains an integer between 1 and 31, which appears to be the "serial number" of the soundbox. It is applied consistently throughout England and on the continent of Europe (we never have the same soundbox turning up at different places at the same time, for example). For the first few months letters occasionally appear; these might be ex-Victor soundboxes, remaining with The Gramophone Company until "standardisation" was completed.

The EMI Archive has ten surviving recording horns, of which I have taken the dimensions. The second column logs the horn(s) used. In this case, the numbers aren't "serial numbers," but "type numbers." Thus, a Type 100 horn would be used for orchestras, or orchestral accompaniments to soloists. Types 11, 11½, and 17 would be for speech, solo vocalists, or small groups of vocalists. Type 11½A was used for solo violin or viola; the surviving example is not a "right cone," but one with its mouth at an angle, and the suspension hook suggests it "looked down" onto the bridge of such an instrument. One Type 01 horn would be used for piano accompaniments, or a pair of Type 01 horns for solo classical piano; here, a surviving Type 01 has an angle in its axis, and was evidently suspended so it "looked down" onto the sounding-board of a grand piano. Horns would be combined for more complex sessions up to a maximum of four. Two of the surviving horns lack type numbers, and there are four types in the Artists' Sheets which do not match numbered ones in the collection. Standard horns do not seem to have reached all HMV's continental studios until the end of 1921.

The third column appears to describe the parallel-sided tube; in section 12.16 we shall see this was where several horns might be combined. Its effect upon the recorded sound is less easy to decide, because the whole purpose of the sheets was to ensure consistency from session to session. I found it impossible to find recordings of the same music recorded with the same soundbox and horn(s) in the same recording-room on the same date with only this factor changed. I found eight examples of *different* music which fitted the other requirements, so I prepared a CD with the different music side-by-side on the same track for comparison by expert listeners.

Unfortunately, it proved impossible to get consistent responses; but I wish to thank the listeners for the time they gave. A further test of self-consistency appears at this point. The code number for this artefact *always* changed when the horns increased from two in number to three (for example, when an extra soloist was featured). We shall see why in section 12.16; but in the meantime the purpose of the extra horn can be deduced from the enumerated artists, and correlates correctly with the type-number of the third horn.

The Artists' Sheets provided me with a major breakthrough, because we appear to know the equipment used to make many recordings, and we can check if its performance can be compensated by our equalisation theories. We can also check the inverse procedure, equalising first and seeing if we can tell what equipment was used. If this proves to be successful, we might then extrapolate our techniques to records made by unknown equipment.

The next sections will describe four parts of the recording machine which affected the sound that was recorded - the mouth of the horn, the air in the horn itself, the parallel-sided tube, and the recording soundbox. To a first approximation, it will be satisfactory to assume that the pattern of the sound wave in the groove was the linear sum of these four effects. If it wasn't, various forms of harmonic or intermodulation distortion would probably have become apparent; in practice such distortions were at an insignificant level. The most significant error to this assumption will be where mismatches occurred at the interconnection of the individual parts. Additional deviations in the overall frequency response would arise if the parts mismatched at certain frequencies.

12.7 The Recording horn - why horns were used

The "megaphone" quality of acoustic recordings is due to the use of a horn. Horns were used because they have the property of matching the impedance of a relatively heavy vibrating diaphragm to that of a lighter medium such as air (see section 12.14 for an explanation of acoustic impedance). They can do this in both directions - whether used for recording sound or reproducing it - but most of the studies have been done for the latter purpose. A theorem known as the "reciprocity theorem" (Ref. 2) suggests this doesn't matter too much. For most purposes, it is possible to take the work for sound reproduction, and use it to help us understand obsolete sound recording machinery - and much of my work begins from this assumption.

The evidence provided by photographs of commercial acoustic recording sessions suggests that about ninety percent featured a single conical horn between six and eighteen inches in diameter at the larger end (150 to 450mm), with lengths between three and six feet (one and two metres). The horns used on "domestic" recording phonographs for amateur use tended to be about half these sizes. Other evidence suggests that multiple horns, in particular, were commoner than the photographic evidence suggests; but this is a complication I shall leave until section 12.15. In an experiment carried out by the author, the sensitivity of a diaphragm to male speech one foot away (300mm) was improved by about 22 decibels when a single conical horn was used.

The acoustic properties of conical horns have not been analysed to a very advanced degree, because it was shown in the mid-1920s that the "exponential" shape gave better results for playback, and research tended to concentrate on that type of horn. Conical horns suffer two significant effects which overlap in a way which makes it difficult to separate them.

12.8 The lack of bass with horn recording

All horns transmit low frequencies with reduced efficiency unless they are made impracticably large, and this has been known from time immemorial. A good mathematical summary for the conical shape may be found in Crandall (Ref. 3). This analysis makes a number of simplifying assumptions, but shows the bass cut-off is twice as powerful as you'd get with the average "tone control." To express the matter in engineering language, the effect is of two slopes of 6dB/octave, the two turnover frequencies being separated by exactly one octave.

Because this shape does not occur elsewhere in analogue sound recording, an approach is to quantify the effect of the higher-frequency turnover. When this is done, the "-3dB point" (see fig. 2 in section 6.7) conversationally becomes a "-4dB point". (It's actually 4.515dB). Both theory and practice suggest that the exact frequency of the "-4dB point" depend on the wavelengths concerned. Although he does not calculate an example, a graph drawn by Crandall suggests a horn with a mouth diameter of 600mm will have a "-4dB point" at around 730Hz. Some second-order effects (such as the proportion of cone which has become a "frustrum" under Euclid's terminology) will affect this turnover frequency slightly, but not the shape of the bass-cut.

Exact measurements of bass-cuts of practical conical horns are difficult, because another effect happens at the same time, which I shall now describe.

12.9 Resonances of air within the horn

The length of a conical horn is taken into consideration in an analysis by Olson (Ref. 4). The effect is to superimpose peaks and troughs at harmonic intervals on the overall bass-cutting shape. To state the matter in words, it is because sounds entering the mouth of the horn are reflected back from the narrow end, and again where the mouth meets the open air, so the air in the horn resonates, exactly as it would in an orchestral brass instrument. (Except, of course, that there is no embouchure to sustain the resonances). I shall call this effect "harmonic resonances" for short.

The pitches of these resonances depend on the length of the horn, so we may need to know how long the horn was before we can compensate it. I have no historical basis for what I am about to say, but it seems recording horns were made with a semi-included angle between eight and eleven degrees, because empirical trials gave harmonic resonances compensating for the bass-cut due to the mouth. But to restore the original sound electronically, we need to separate the two processes, because they have quite different causes and therefore cures.

12.10 Experimental methodology

Unfortunately, it is not possible to measure the first effect, because of the second effect. There is no way of dampening the harmonic resonances of a conical horn, except by "putting a sock in it," and then we cannot measure the exact bass-cut, because the treble is now muffled! This suggests several approaches, and currently there is no consensus which is best.

(1) Quantify the bass-cut by listening to acoustic recordings reproduced to a constant-velocity characteristic (section 6.4), and set a bass-compensator empirically. Conducted by a large number of practised listeners, on subject matter known to have been made with a single horn of known mouth-size, should give results consistent within a couple of decibels.

(2) Rig an acoustic recording-horn (which must be terminated by an acoustic recording soundbox, or the harmonic resonances will have inappropriate magnitudes), perform into it and simultaneously into a nearby high-fidelity microphone, and adjust both the bass-cut and the anti-resonant circuitry until the two sound the same. This has the practical disadvantage that sound is reflected back *out of* the horn, and the microphone must ignore this sound.

(3) Observe the impulse response of a real horn (for example its response to an electric spark), neutralise the resonances as displayed on an oscilloscope, and then measure the remaining bass-cut.

Brock-Nannestad was the first to carry out measurements of the axial response of one of the surviving acoustic recording horns at EMI Hayes (Ref. 5). These measurements included all the effects of the horn piled on top of each other, but both the bass-cut and the harmonic resonances are visible in Brock-Nannestad's graphs.

At the University of Cambridge, Paul Spencer undertook seminal research for his thesis "System Identification with application to the restoration of archived gramophone recordings" (Ref. 6). This included theoretical studies of the harmonic resonances of a horn, and he tested his theory with measurements upon a small phonograph horn so a computer might know what to look for (to oversimplify drastically). He then wrote a computer program to analyse a digital transfer of an acoustic recording. He confirmed his process was accurate when the computer successfully found the resonances after a recording of a BBC announcer had been made through the same horn.

The main part of the paper outlined two mathematical techniques for determining the spectral distribution of the resonances in an ancient recording, which in turn tells us the length of the horn - the "System Identification" part of the problem. Then, the computer can be made to equalise the resonances - the "Restoration" part of the problem.

Unfortunately the program was never marketed; but the first point of significance is that the mathematical process gives results of an objective rather than a subjective character. So it has more validity to the archivist, and can be repeated and (if necessary) reversed at a later date if the need arises.

Another significance is that Spencer's work gave the first unchallenged evidence of when it is *inappropriate* to neutralise the resonances of a horn. Spencer specifically tried his process on a 1911 Billy Williams record, which comprised speech, instrumental music, and vocal music recorded during the same "take," recorded through a horn with conspicuous metallic vibrations. This was supposed to check the process didn't give different answers on different subject matter when the horn was known to have been the same. The process indeed gave consistent results; but Spencer found resonances suggesting there had been *two horns* of different dimensions - one for the speaker/vocalist and one for the band. While it might be possible to crossfade from one to the other for certain passages, it would be difficult to achieve correct equalisation when the singer and the band were performing together.

This particular recording was also chosen because it was "noisy" – audio historians may appreciate the sentiment when I say it was a First World War "Regal" laminate! Spencer's thesis makes it clear that the identification of the resonances was considerably less precise, but the presence of two horns was quite unambiguous.

12.11 Design of an Acoustic Horn Equaliser

Using the theories of Crandall, Olson, and Spencer as a basis, I designed an analogue equaliser which would do the same job at a fraction of the cost. I am very grateful to Hugh Mash for constructing the first prototype, and I am grateful to Adrian Tuddenham for two subsequent versions (based on different acoustic models). At present, subjective judgements show that all three have promise. If an ideal circuit emerges before this goes to press, I shall supply an appropriate circuit-diagram. However, when used on actual historic recordings, we have learnt that analogue restoration operators must have continuously adjustable controls to "tune" and neutralise the "honk" of a practical horn.

In experimental work, test-tones, live speech and suitable music were played into a horn. The theoretical works by Crandall, Olson and Spencer were confirmed in broad outline; but work is still in progress to measure the exact *strength* of the bass-cut and of the harmonic resonances. There are considerable experimental difficulties, some of which are not fully covered by the theories.

12.12 Resonances of the horn itself

This is a section for which I shall *not* be providing recipes for compensation purposes - for two reasons. Firstly, we cannot envisage any electronic or mechanical way of solving the difficulty, and secondly there are ethical reasons why we shouldn't anyway.

The mechanical resonances of the metal of the horn might theoretically be neutralised with analogue electrical networks as well; but it would require circuitry with unprecedented precision. It would also require the invention of a new technique to determine the values of hundreds of separate resonant elements whose features overlap. And all this cannot be done unless we know which horn was actually used on each recording, and its precise mechanical characteristics are known in great detail and with unparalleled accuracy.

Most horns used by recording experts seem to have been made of tinplate. With one exception, the ones surviving at Hayes have a thickness of 25 thou (0.635mm). (The exception is made from ceramic material an inch thick). We think the experts selected such horns partly so they would reinforce the music with resonances of their own (a technique re-invented thirty years later, when the steel "reverberation plate" provided artificial reverberation instead of a soundproof "echo-chamber."). It is also known that the decay-time of a horn's resonances could be controlled by wrapping damping tape around it, and several horns would be provided with differing amounts of tape so that rapid comparisons could be made by means of trial recordings. (Ref. 7).

In effect, the horn functioned somewhat like a bell, having transverse waves moving circumferentially around the metal; but unlike a bell, a conical horn resonated at a wide spectrum of frequencies, not a few fixed ones. Unhappily none of the damping tapes have survived. They are only depicted in photographs. It is not even certain what they were made of, nor how they were fixed to the horn.

Most of the surviving metal horns at Hayes have about a dozen small holes about one-sixteenth of an inch in diameter. It is possible they were for anchoring the tapes, but there are at least two other explanations. Railway engineers who fitted carriages with the first steel wheels found they made a terrible shrieking noise (the wheels, I mean). The cure was to drill small holes through the face of each wheel to break up the tendency to resonate. And I am very grateful to Sean Davies for drawing my attention to a memorandum about Victor's recording processes, written by the Gramophone Company's expert Fred Gaisberg after his visit to the USA in 1907. He reported "The Horns are made of Block Tin with very few perforations for ventilating." My experiments have been done with horns with no holes; but I found it easy to "pop" the diaphragm when speaking into the horn and uttering a "p" sound. (Much the same can occur with microphones today). It is possible small holes allowed the blast to escape without too much sound energy leaking away.

So we have two major difficulties in dealing with the metal of the recording horn. First, ethical considerations which arise from neutralising conscious choices made by the original experts, and second the difficulties of analysing such behaviour using conventional calculus or equalising it using conventional equalisers.

But digital processing was found to offer a route towards a solution of the latter problem at a surprisingly early date. In 1974 Stockham (Ref. 8) introduced his process for the Victor records made by Caruso, which had remained best-sellers more than half a century after the tenor's death in 1921.

It should perhaps be explained that the term "convolution" means subjecting a digital data-stream to a numerical process in which the digits are handled in the time-domain - that is, not necessarily in chronological order. Because *both* the wanted signal and the nature of the distortion are unknown in acoustic recordings, Stockham called his process "Blind Deconvolution," and this phrase has now entered the vocabulary of signal-processing researchers. To oversimplify greatly, Stockham took a modern recording of the same piece of music (assumed to have been made with a flat frequency response), and used it as a "prototype" so an older recording could be matched to it. At the time his paper was written the bandwidth was limited to 5kHz (possibly to reduce the costs of the computer processing, which took two hours to process a four-minute recording using a Digital Equipment Corporation PDP-10 mainframe). Judging from the graphs in his paper, he divided this frequency range into 512 bands.

Stockham's pioneering work involved comparing Caruso's 1907 recording of "Vesti la giubba" with Björling's modern version of the same aria. An experienced sound operator would immediately see several defects in this idea. Firstly, there is no such thing as a "flat" recording of such a piece of music - it could only be approached, and never achieved, using a sound calibration microphone in an anechoic chamber. Even the best recording studios offer chances for many multipath routes between sound source and microphone, giving narrow frequency-bands with complete cancellation of sound. When convolved from prototype to processed recording, these would become peaks with amplitudes far beyond those discovered by Brock-Nannestad. Secondly, the voices of Björling and Caruso differ because of the different dimensions of their nasal, mouth, and throat passages, which cause a specific distribution of resonances to be superimposed on the sounds of the vocal cords. Humans use these resonances to tell one person from another. There is a risk that this process would make Caruso sound like Björling. Thirdly, Caruso had a wind band accompanying him, Björling a conventional orchestra. The effect of this is unimaginable. Stockham reduced these difficulties by "smoothing" the

characteristics of his prototype, so the shorter resonances were ignored while the longer resonances were treated.

By comparing the smoothed prototype with the 1907 version, Stockham obtained a very spiky response curve which was supposed to indicate the response of the acoustic recording machine. When this was inverted and applied to the acoustic recording, many effects of the horn certainly disappeared. The result was issued on a published L.P disc (Ref. 9), and created a great deal of interest among collectors of vocal records, without any consensus about the fidelity of the technique emerging. One defect which everybody noticed was that the raw material supplied to Stockham frequently suffered from clicks and pops, which the process turned into pings and clangs of distracting pitches.

Stockham continued to process RCA's Caruso recordings for many years, presumably making improvements as he went along, although details have not been published. For their part, RCA supplied transfers from their surviving metalwork. Eventually all the band-accompanied and orchestrally-accompanied published recordings were processed. (This immediately rings alarm-bells to a restoration operator. Does the process fail on recordings with piano accompaniment, or is it simply that a piano-accompanied prototype must be available?). The final corpus was issued in 1990. (Ref. 10).

So far, Stockham's process has not been applied to recordings made with horns with known properties, so we cannot judge how faithful his process is. But the technique is very powerful, and although my ethical reservations remain, it is the only possible way to cure resonances in the metal of the horn.

12.13 Positions of artists in relation to a conical horn

Obviously, the closer an artist was to the recording horn, the louder he would seem to be. One writer, who was admittedly extolling the advantages of electrical recording, stated that a tolerance of one inch was aimed for. Frankly, I doubt this; if so, recording experts would have tied their singers to a stake! And Gaisberg makes it clear the Victor Orchestra had more freedom than that. However, the balance between two or more artists, and the balance between soloist and accompaniment, would only have been settled by distances from the horn(s). Small movements by each party might accumulate, and imbalances would then double.

It has widely been reported that the quality of the sound also varied with distance. Some collectors have puzzled why the famous Melba "Distance Test" recording was made in London in May 1910, especially since the matrix number suggests it was the last recording of the session! (Ref. 11). Brock-Nannestad did spectral analyses of the four vocal passages from this recording (Ref. 12), without publishing any concrete conclusion; but in Reference 1 he showed the recording was an attempt by the Gramophone Company of London to mollify the US Victor company, which wished to publish these recordings in America.

My own speech tests with a replica horn suggested little variation with distance, except that when the source of sound was very close I had bigger pressure differences between the inside and the outside of the horn, so the metal of the horn itself vibrated more. There is also the consideration that the closer I was, the more my face reflected sound into the horn than the open air did. But any differences are small, and typical of the clues listeners use to establish the "perspective" of artists (clues such as the ratio of direct

sound to reverberant sound from the studio walls). I see no need to delete these clues, since I consider they are part of the original sound.

One major effect which remains unresearched, and which I suspect will be impossible to solve, is that the mouth of the horn was so large that diffraction effects occurred when the artist was off-axis. The theoretical work has assumed axial pickup so far; but it is well known that high frequencies will be attenuated if they are emitted off-axis. Indeed, Reference 7 shows a photograph of a session with the brass placed to one side specifically to muffle them somewhat. Thus it should be remembered that electronic equalisation can never be made exact for higher frequencies (above about 500Hz) performed off-axis. The only cure would have been two or more horns pointing in different directions and connected at their necks, and this may explain why there is an imbalance between the photographic and other evidence: It was regarded as a trade secret. Other horns would be unplugged when the photograph was taken; yet there is ample evidence that multiple horns were fairly common. Which brings us to the next piece of hardware.

12.14 Acoustic impedances

Before I continue, I'd better explain a scientific concept which isn't at all easy - the concept of "acoustic impedance."

When an electrical voltage is applied to a wire, the quantity of the resulting electric current depends upon resistance of the wire. When alternating voltages are applied to electronic components, the resulting current may not only depend on the resistance, but may also vary with the frequency of the applied voltage. To make the distinction clearer, engineers call the latter phenomenon "impedance" rather than "resistance." It is precisely because electronic components exhibit different impedances at different frequencies that we are able to build electronic filters to emulate acoustical and mechanical phenomena. When sound energy flows down a tube, it too has to overcome the impedance of the tube - its "acoustic impedance."

We may lose energy if we do not take impedances into account. Ideally, impedances should be "matched" whenever energy flows from one component to the next. This applies to electrical energy, acoustic energy, or the energy of vibrating mechanisms; and it even applies when we are converting one form of energy into another.

For early sound recording experts without amplifiers, this was essential. The performers emitted only a few watts of acoustic energy, and the mechanism collected only a small proportion of this; yet it had to be sufficient to vibrate a cutting tool. A horn was used to match the acoustic impedance of air - quite low - to that of something comparatively high - the recording diaphragm.

But to collect sound from a number of performers, a number of horns were needed. I have no doubt that early experimenters tried connecting two horns to a recording machine by means of a Y-shaped piece of rubber tubing. Yet this did not double the loudness of the recordings; on the contrary, it made matters worse. In plain English, half the sound picked up by one horn would then escape through the other horn.

In practice things were even worse than this, because the recording diaphragm could not be matched to the first horn equally well at all frequencies. The acoustic energy had to go somewhere; if it wasn't vibrating the cutter, it could only either be absorbed - turned into heat - or reflected back; so considerably more than half the sound energy

might escape through the second horn. An exception occurred when the sound entering the two horns was very similar (the technical term is "highly coherent.") In this situation, the sound pressures and rarefactions at the necks of the two horns were near to or at synchronism, so little or no sound escaped. Something like this occurred on HMV solo pianoforte records made with two Type 01 horns. Although the horns were situated over different parts of the piano, much of the sound from the piano's sounding-board would be coherent, so leakages would have less effect.

When there was no coherence, there might be a third reason for the faintness of the actual recording. Each horn had its own impedance characteristic. We saw earlier that horns had longitudinal acoustic resonances depending on their lengths, and their acoustic impedances were highest at resonance. (A trumpet player uses this effect to match his embouchure to get a certain harmonic from his instrument). If the two horns were the same length, the two acoustic impedances would have their maxima at the same frequencies, and the sound energy would indeed be divided by two. But if the horns were different, their acoustic impedance characteristics would be different. While one horn was resonating and strengthening the sound, the other would have a lower acoustic impedance, so considerably more than half the energy from the first horn would escape through the other.

Earlier I described some of the horns surviving at the EMI Archives at Hayes. There are many shapes and sizes, and when we calculate their characteristics, we find they never exactly coincide. So how did acoustic recording experts manage as well as they did?

Electrical effects are often excellent analogues of mechanical or acoustic phenomena. Many readers will remember similar problems when connecting two microphones in the early days of tape recording, when you couldn't get an electronic sound-mixer for love or money. How did you stop one microphone from short-circuiting the other?

12.15 Joining two horns

A few years ago, Mr. Eliot Levin (of Symposium Records) donated to the British Library Sound Archive a small collection of acoustic recording artefacts, such as diaphragms and soundboxes. There is one Y-shaped connector fashioned from three pieces of laboratory rubber tubing joined together with rubber solution. It is very brittle, being some seventy years old; but by shining light down it, you can just see that one arm of the Y is partially closed off. I conjecture this arm would have been connected to a horn for a soloist, while the other went to a horn for picking up the accompaniment. Thus the accompaniment would be conveyed to the cutter with reduced leakage, at the expense of the soloist being quietened by a few decibels.

This exactly mirrors the early days of amateur tape recording. If you were balancing a soloist in front of (say) an orchestra, you would have one microphone for the soloist, and another picking up the overall orchestral sound. Because the soloist was close to his/her microphone, that volume control did not have to be fully up, and the resistance of the control itself stopped the orchestral mike from being short-circuited.

In section 12.12, I mentioned how Fred Gaisberg (of the London-based Gramophone Company) visited the US Victor Company in 1907 to study their methods. His memorandum contained this hitherto-puzzling sentence: "Piano accompaniments are made by using the ordinary Y with a small hole such as we now use for Orchestra accompaniments." I think that sentence is no longer puzzling.

12.16 Joining three or more horns

So far as the record-buying public was concerned, the big breakthrough happened at Victor in 1906. In his book *The Fabulous Phonograph*, Roland Gelatt observes that Caruso's recordings made that February were not only his first to have an orchestra, but they were also slightly louder and more forward-sounding than anything achieved previously. To my mind, the significant phrase is "more forward-sounding." Although the acoustic recording process modified the original sound very considerably, the distance of the artist from the apparatus is usually abundantly clear! The only practical method of getting a "more forward sound" was to use several horns with artists placed close to each of them. Fred Gaisberg's memorandum depicts three horns in use at once, plus the secret of how the horns were connected. (Actually, I must be honest and say it is only one of several possible technical solutions, and that I quite expect a supporter of Edison technology to tell me a solution was found at Edison's studio first).

It seems that all Victor horns of the time (and of the Gramophone Company in Europe) ended with a short section of tube 11/16ths inch (17.5mm) inner diameter and 3/4 inch (19mm) outer diameter. Connections were made by fitting such sections together with short pieces of rubber tube over the joins. Gaisberg's notes show a gadget for combining three horns to record "The Victor Orchestra." Basically it consisted of a straight piece of tube of the same diameter as the end of the horn. (The exact length wasn't given; it was apparently four or five inches). It is clear that the main orchestral horn would be connected to one end, and the recording soundbox to the other. But this straight tube was modified by two more tubes entering it from two different directions, with bores tapered like a sharpened pencil. The conical shapes formed natural extensions to the narrow ends of the two extra horns, with the diameter at the narrow end of the taper being 3/16ths of an inch (4.75 mm). Sound waves from the main horn were unobstructed, but individual soloists could be brought close to the additional horns and their sounds injected into the main tube. The leakage from the main horn would have been less than one decibel (hardly noticeable), and because the hollow conical extensions accurately matched the additional horns, losses due to poor matching were smaller.

Gaisberg's memo also shows how the musicians were located to give the correct musical balance, to counteract the horns not being equally sensitive. The memo also deals with the layout for sessions featuring vocal soloists when they were accompanied by the Victor Orchestra. Here a different gadget was used, because the soloist (who needed the unobstructed horn) was to one side. There is no detailed drawing of that gadget.

In section 12.6 I mentioned how HMV documented the coupling-gadgets used for each recording. We now have an explanation - not a proof - for why the identification *always* changed when the horns increased from two to three. For two horns a Y-tube might be used, for three a gadget. I also thought the identification was noted because the gadgets influenced the sound quality. My listening-tests showed that the sound quality was indeed influenced, but only by a small amount. I now think the gadgets' numbers were entered because they provided an extremely concise way of logging the layout of the musicians.

12.17 Another way of connecting several horns

There is another way to use two horns whilst minimising leakages. This is to connect one horn to one side of the diaphragm, and the other horn to the other. Paul Whiteman, the

dance-band leader who recorded for Victor from 1921 onwards, later gave his recollections of acoustic recording days (Ref. 13). Instead of the recording-machine being placed behind a wall or vertical partition at one end of the studio, he describes it hidden inside a four-sided box, with what looked like ladders on each side, erected in the middle of the studio. The four walls of the box each had a recording-horn protruding some five feet above the floor ("in the form of a four-leaved clover"), and the recording-expert was encased with his machine so no-one could see what he was up to.

In this context, it seems the four horns fed the recording-machine by the shortest possible routes, namely through a pair of Y-tubes, each to a different side of the diaphragm. The only apparent alternative would have been a complex array of pipes all terminating on one side of the diaphragm. Not only would this have been less pure acoustically, but why entomb the recording-expert?

The electrical equivalent of a double-sided diaphragm would be two unidirectional microphones on one stand pointing in opposite directions, or one bi-directional microphone (which comes to the same acoustically). Here it would be necessary to move either the microphone-stand or the soloist to get the right musical balance, so this idea wasn't used by early tape enthusiasts very often; but again, no electrical energy is lost, and I myself used a similar technique on several commercial recordings before I got my first stereo mixer.

As a result of my writings, I am very grateful to George Brock-Nannestad for effectively confirming my theory. He wrote (Ref. 14) that Victor had invented an improvement which they called "the DR System", in which two recording soundboxes were coupled together at their centres by a steel wire under tension. The effect of this would be very similar to one soundbox addressed from both sides, while permitting as many as eight horns.

12.18 Electrical equalisation of recordings made with parallel-sided tubes

How may these researches be applied to the task of reversing the effects of the recording equipment?

The air in a parallel-sided tube has harmonic resonances rather like those in a horn, and matching at each end is similarly important. But what happens when a single conical horn is joined to a parallel-sided tube? Do the resonances of the air in the horn and in the tube continue to exist separately? Or does the air in the combination behave like one body of air? Or are there elements of both of these? I am not aware this problem has been handled mathematically; most published studies were done with exponential horns after acoustic recordings ceased to be made. Would it depend on the semi-included angle of the horn - in other words, the taper by which the horn differed from the tube?

The question can be answered experimentally. Listening trials with a full-sized tube and a replica conical horn are quite unambiguous. The air in the horn and the parallel-sided tube each have their own harmonic resonances in approximately equal proportions. Thus any quality side-effects of the parallel-sided tube may be compensated - so long as two conditions are met. First, the resonances are audible or measurable; but as I indicated in section 12.3, I did not get consistent results in A/B listening-tests. Second, it is assumed the parallel-sided tubes were not imposed for purely subjective reasons, in which case it would be unethical to reverse the experts' work.

The HMV Artists' Sheets show one case (the Catterall String Quartet) where the coupling artefacts were changed from "9" to "33" as the music changed from Beethoven

to Brahms. That was on 18th June 1923; and the following day the exact reverse occurred as they completed the Brahms and started some more Beethoven. The horns remained constant throughout all this (four "Type 60"s). The two Beethoven quartets were from his early "classical" period (Opus 18, Nos. 1 and 2), and since the Brahms was from the height of the "romantic" tradition (Opus 51 No. 1), a definite musical effect may well have been sought. My subjective judgement is that the Brahms has more "warmth" in the octave 125Hz - 250Hz.

Although we know which tubes were used for HMV recordings in the years 1921-1925, we have no knowledge of their dimensions. For all other acoustic recordings, we have absolutely no historical knowledge at all! Thus my current view is that we should put the parallel-sided tubes (if any) "on the back burner" until future methods of analysing recordings enable us to quantify their effects.

In the meantime, the following observation is offered. When I made the tube for the above-mentioned experiment, I deliberately chose a length in which the tube's harmonic resonances would be low-pitched enough to be heard by a 1920's recording expert, yet quite different from those of the horn so they would be easier to hear. Thus one set of resonances filled in the frequencies between the other set. The overall sound was noticeably truer to the original; but there was a drop in overall volume (about six or eight decibels).

12.19 Electrical equalisation of recordings made with multiple recording horns

The next point is to decide whether a particular recording was actually made with two or more different horns. (If they were all the same, the horn-quality cure can fix them all at once). For certain types of subject matter (mainly solo speech and amateur phonograph cylinders), only one horn would be used. For HMV records 1921-1925 we can look in the Artists' Sheets. But in the majority of cases, we can only "try it and see." Once the cure has dealt with the principal sound, does other "hornyness" remain?

If so, we must first decide whether to take the matter further. All three methods of coupling (the Y-shaped connector, the parallel-sided tube, and the double-sided diaphragm) effectively separate the additional horn(s) from the main horn, so the main horn continued to operate as if the additional horn(s) weren't there. Thus it is always perfectly valid to have the cure for the main horn in circuit. If it cures the sound of the principal performer, that may be sufficient for the intended application. (The accompaniment may be considered a convention of the recording studio, rather than representing real life). But this may not be true of other music. A band or orchestra with many parts of equal importance, a concerto, or a "concerted" session (this is Gramophone Company terminology for several vocalists), may tempt one to correct the additional horn(s) as well.

My first technique was to copy the unfiltered recording to one track of a multitrack tape-recorder. I then "track-bounced" this track through my so-called "hornyness-cure" to several other tracks, each with appropriate settings. This kept the several tracks in phase, and they could then be mixed together depending on which horn(s) were in use at which time. The current practice is to have several hornyness-cures with their inputs connected in parallel, and their outputs routed to different faders on an electronic mixer.

The hornyness-cure often "colours" the background noise, so sometimes you can hear the surface-noise change distractingly whenever you concentrate on getting the

foreground sounds right. Fortunately, most horn resonance frequencies have good separation from surface-noise frequencies, and most horn resonances can be removed without side-effects. But one sometimes has to make a subjective compromise, either by leaving the tracks at a fixed setting, or by changing the balance very gradually.

This may give an artistically preferable result; but unfortunately I cannot think of a rigorous archival way of neutralising the effects of different horns used simultaneously. Even without the noise problem, we have to assess the contributions of the different horns by ear. Subjective judgement always seems to be needed, and we shall need another paradigm shift before we can be sure of restoring all the original sounds in an objective manner.

12.20 The recording soundbox

In the next few sections, I hope to show how the recording soundbox and its diaphragm affected the sound we hear from acoustically-recorded grooves today. I use the phrase "recording soundbox" to mean the artefact which changed acoustic energy into the mechanical vibration of a cutting-stylus. It may not be a very appropriate word, because it wasn't always like a box; but I borrow the term from its equivalent for reproduction purposes. I shall assume readers are familiar with the latter when I discuss the differences between them.

But I must start with another short lecture on a further aspect of physical science, so let's get that over with first.

12.21 "Lumped" and "distributed" components

Electrical recording was a quantum leap forward in the mid-1920s thanks to an idea by Maxfield and Harrison of the Bell Telephone Laboratories. Methods had been developed for electrically filtering speech sounds and transmitting them over long telephone cables. The two scientists realised equations for designing electrical filters could be applied to mechanical and acoustical problems by taking advantage of some analogous properties of the different bits and pieces. An electrical capacitor, for example, could be regarded as analogous to a mechanical spring, or to the springiness of air in a confined space. Using such equations to design mechanical and acoustical devices on paper saved much trial-and-error with physical prototypes. The analogies soon became very familiar to scientists, who switched rapidly between electrical and acoustic concepts in their everyday conversations, to the utter bewilderment of bystanders.

The principles were used for the design of an electrical cutting-head and an acoustic gramophone, which were marketed in America in 1925. They were fully explained in Percy Wilson and G. W. Webb's book "Modern Gramophones and Electrical Reproducers" (Ref. 15), which naturally concentrated on the issues affecting reproduction. But the same principles continued to help recording until at least 1931, when Western Electric designed a new type of microphone (section 6.17 and Ref. 16). This was the first electromagnetic microphone to have a reasonably wide flat frequency response with low background noise - in fact, the first of a long list of such mikes which appeared over the next half-century. It was surprisingly complicated to design, and in my view it was the most successful application of the principles of analogies.

However, these analogies all depended upon a hidden assumption - that a pure electrical capacitance would be exactly equivalent to a mechanical spring or a trapped volume of air. This was an oversimplification of course. You cannot have a mechanical spring or a trapped volume of air without also having mass and friction (amongst other things). So the theory, which started from the behaviour of idealised electronic parts, could only be applied to "lumped components" - springs which had no mass, for example, or whose mass could be assumed to be concentrated at just one point.

Maxfield and Harrison's playback soundbox differed from pre-1925 soundboxes in two ways, the second somewhat counter-intuitive. Firstly, they made the diaphragm of light stiff aluminium, which was domed and corrugated to make its centre substantially rigid. (This then behaved like a "lumped component.") Secondly, they put a spring between the stylus-lever and the diaphragm - the "spider." Starting from page 69, Wilson and Webb's book took many pages explaining "lumped component" principles to show why a spring between the stylus-lever and the diaphragm gave better high-frequency reproduction. The idea seemed insane at the time.

By 1932 the lumped component analogy was at its height, and when the engineers at RCA Victor decided to reissue the acoustic recordings of Caruso with new electrically-recorded accompaniment, they assumed the resonance of the acoustic recording soundbox could be neutralised electronically. Such a resonance would have emphasised notes of a narrow range of frequencies between 2 and 3kHz, which might have accounted for the "tinnyness" of the sound. But they were wrong; parts of the soundbox behaved like "distributed components." The springiness had been distributed throughout the material of the diaphragm and its surrounding rubber, so "lumped component" analysis simply didn't work, and it wasn't possible to recover Caruso's voice to suit the new electrically-recorded accompaniment.

On the contrary, listening often suggested notes in this region didn't get recorded at all. Acoustic recording-experts had avoided these situations whenever they could, of course; but I can think of several examples from my own collection, and I'm sure you can as well (Ref. 17). By working out the frequencies of the missing notes, it can be shown that (in Britain, anyway) the missing frequencies gradually became higher and higher over the years, varying from 2343Hz in 1911 to 2953Hz in 1924.

More recently, at least three different workers - I will spare you their names - have proposed identifying the acoustic recording diaphragm resonance by analysing the "coloured" background-noise, either because the diaphragm was stimulated into its own vibration by the wax blank, or because it was picking up the hiss of the vacuum pipe nearby. These results have always proved inconclusive. I am afraid there has never been a conspicuous diaphragm resonance we can latch onto - although this doesn't mean the background-noise is useless, as we shall see in section 12.27.

12.22 How pre-1925 soundboxes worked

So how did earlier soundboxes work - whether for reproduction or recording? I have been searching for a complete account without success. Lord Rayleigh's book "The Theory of Sound" (Ref. 18) goes into the theory of circular diaphragms with distributed properties, but develops a model which is oversimplified for our purposes - a uniform circular plate like those in early telephone earpieces. Soundboxes behaved in a much more complex manner. The diaphragms were circular, it is true; but they had mass stiffness and resistance distributed throughout them, they were surrounded with rubber which

provided both shear and longitudinal springiness and resistive damping, the stylus-lever had springiness at its pivot, and it formed a lumped mass where it was attached in the middle.

In general, recording-soundboxes were smaller and more delicate than reproducing soundboxes, so surviving examples always seem to be "in distressed state" (as auctioneers say). Rubber has always perished, diaphragms are often broken, and cutting styli have usually vanished. So we cannot simply measure their behaviour today. But it is clear that recording diaphragms had diameters from 1 3/16 of an inch (30mm) to 2 inches (51mm), these extremes being represented by examples from His Master's Voice and Columbia. Towards the end of the acoustic recording era, they all seem to have been made of flat glass with thicknesses from 7 thou (0.17mm) to 10 thou (0.254mm). The edges were clamped in rubber, of various configurations but essentially coming to the same thing - rectangular cross-section pressing against both sides of the glass. None of them have much springiness at the pivot of the stylus-lever, which therefore feels surprisingly "floppy." (Refs. 19 and 20).

I have been experimenting with a lateral-cutting soundbox of unknown origin intermediate between these two diameters (1 3/4 of an inch, 44mm) to find how it worked. I replaced the perished rubber (whose dimensions were still apparent), but could not be certain the new rubber had exactly the right physical properties. Unlike the HMV and Columbia patterns, this soundbox also had a mechanism for "tuning" the diaphragm, comprising an annular ring in a screw-thread which bore upon the rubber rings. Tightening it would raise the frequency of the diaphragm's "piston-like" mode of resonance.

It is clear that there were two effects at once, one counteracting the other. The classical piston-like resonance was certainly there at the edge of the diaphragm, and it could be tuned by the screw mechanism. But the stylus-lever added mass at the diaphragm's centre. This made it comparatively difficult to move, rather like a small island in the centre of a circular fishpond. Low frequencies caused the whole diaphragm to vibrate normally, as if the whole fishpond was being shaken up and down slowly by an earthquake. At higher frequencies waves were set up in the diaphragm, and a mode could develop in which waves were present at the edge, but with a "node" (no vibration) at the centre, so the stylus was not vibrated. This phenomenon had been described by Lord Rayleigh, and is known as "the first radial mode of breakup."

Thus acoustic recording experts used soundboxes made from certain materials with certain dimensions so the "piston-like" mode of resonance would be counteracted by the first radial mode of breakup. I am certain they did not use "lumped analysis," let alone "distributed analysis," to design their soundboxes! So how did they do it?

Obviously, they used their ears. Probably the best way would have been to say "she sells sea-shells beside the seashore" into the recording machine, trying several soundboxes one after another. Then they would pick the soundbox which recorded the sibilants as clearly as possible without turning them into whistles. To judge this, a playback soundbox might also be needed whose performance was beyond reproach; but this could be tested by listening to the surface-noise, not the recording. A playback soundbox which made the surface-noise sound like a steady hiss, rather than a roar or a whistle, would have been a good candidate. In any case, the human ear can easily judge the *differences* between two soundboxes, eliminating the characteristics of the one which is making the continuous noises.

12.23 Practicalities of acoustic recording diaphragms

Numerous written accounts show that the diaphragm was considered the most critical part of the acoustic recording process. Not only did it have to be assembled so the first radial mode of breakup and the piston-like resonance counteracted each other as closely as possible, but different diaphragms were used for different subject matter. The standard stories say that thick glass was needed for something loud with high frequencies (such as a brass band), whereas thin glass was used for something faint and mellow (such as solo violin). It was the height of the romantic era in music, and mellowness was, if anything, considered an advantage. Thus experts intuitively called upon another scientific principle, the "power-bandwidth product" (section 2.3), which makes it possible to trade high-frequency response for sensitivity.

The balance between the piston-like resonance and the first radial mode of breakup cannot be perfect, because they have different causes, and therefore give different shapes to the frequency response. The piston-like resonance causes a symmetrical peak about 10 or 15 decibels high, while the first radial mode of breakup takes out a comparatively narrow slice to a great depth, and the two sides of the slice aren't the same. Even with laboratory test-gear, I have not been able to decide precisely how to tune my experimental soundbox, because you cannot get a perfect frequency response whatever you do.

Thus each individual recording soundbox still had a sound quality of its own, and we can sometimes hear this on multi-sided sets of orchestral music. (Ref. 21). Straightforward listening suggests two different soundboxes were used during these musical works. Electrical equalisation has so far proved powerless to permit these pieces of music to be joined up seamlessly.

Record companies evidently learnt from examples like these, because consistency became important for multi-sided sets. According to a 1928 article in The Gramophone magazine (Ref. 20), English Columbia used only two soundboxes for the remainder of the acoustic era.

According to the Artists' Sheets for the years 1921-1925, HMV had many more soundboxes (they had studios in many more locations), but the serial number of the diaphragm used on each take was always logged. Analysis also shows that experts travelling to a new location always took two diaphragms from their "base" studio with them. Whenever possible, they recorded each title twice, once with each diaphragm, in case one had changed its sound. But analysis also shows that the *same* diaphragm(s) were nearly always used for *different* subject matter. Thus it seems the idea was to have a large number of diaphragms sounding identical, so Chaliapine might sound the same whether he was recorded in London or Paris. Volume adjustments seem to have been achieved by positioning the artists and selecting horns and coupling gadgets, as we saw earlier, rather than changing the diaphragm. The only times different diaphragms were used was on certain solo piano records. Evidently "romanticism" was considered preferable to a fainter recording!

You will notice that I suddenly switched from saying "soundbox" to "diaphragm" during that last paragraph. This is because HMV didn't use a soundbox with a matching cavity, as described in Wilson & Webb's book. Instead, the end of the horn terminated in a parallel-sided tube cut precisely at right-angles and clamped firmly to the recording machine. A diaphragm was pivoted close to the tube less than a hundredth of an inch away, so the *diaphragm* was free to move up and down with the level of the wax,

without the horn having to move. There was therefore no matching cavity, and it was much easier to change diaphragms.

12.24 The rest of the soundbox

We now study the remainder of the soundbox, namely the pivot, the stylus-lever, and the stylus. To minimise the mass, a tiny cutting stylus was attached directly to the stylus-lever with sealing-wax. It therefore seems logical to consider the latter two together.

Brock-Nannestad's paper (Ref. 5) shows two patterns of soundbox used by the Victor/Gramophone companies, with quite different pivot mechanisms. However, I consider it is too early to worry about the performance of these. "Distributed-component" analysis clearly suggests that both types would have vibrated exactly like "lumped components" at frequencies below 5kHz, and we shall have quite enough difficulties restoring these frequencies anyway. Perhaps studying the two types may be critical for recovering the original sound with fidelity one day, but frankly I am doubtful. Even with the best modern technology, such high frequencies are usually drowned in background noise. (This doesn't mean high frequencies aren't audible, only that we must rely on psychoacoustic knowledge to get them out). And lumped-component analysis suggests even-higher frequencies would be attenuated rather than boosted, so matters can only get worse the higher we reach.

It is also clear the wax would have imposed a load on the stylus-lever, but it seems the load would be "resistive" in character (that is, the same at all frequencies). Its effect would have been to dampen the piston-like mode of resonance. It would have varied inversely with the recorded diameter on a disc record. Since the piston-like mode and the first radial mode of breakup were designed to offset each other, there is little audible difference between the outside edge of a disc and the inside. Thus we may consider the load of the wax to be almost insignificant - compared with the other effects, anyway.

During this chapter we have studied all the parts of the hypothetical acoustic recording-machine I described in section 12.3, and examined how they affected the wanted sound. I am sorry we end in a down-beat fashion. We do not actually know how to equalise the performance of a diaphragm!

There are three reasons for this. First, we often do not know the frequency at which the first radial mode of breakup took place. Modern spectral analysis may sometimes help, but as I said earlier, recording-experts often dodged musical notes in this region, and the evidence simply may not exist. With the aid of the HMV Artists' Sheets, we may sometimes identify the actual diaphragm, and get the information from other recordings made with the same kit on nearby dates; but then comes the second difficulty. We would need to boost the attenuated frequencies, and we would get a highly coloured peak in the surface-noise. (Or we might learn how to synthesise the missing sound one day). Thirdly, we do not know precisely how the breakup was judged to compensate for the piston-like mode of resonance, so we have no idea of the effects of the main resonance on either side of the gap. At present, we can only leave the effects well alone.

12.25 Notes on variations

In section 12.3 I outlined a "typical" acoustic recording-machine, and promised to deal with some of the "variations" later. So here we go.

The first point is that many early cylinder and disc records were not made with recording horns at all, but with speaking-tubes. To judge from pictures of Berliner's and Bettini's machines in Roland Gelatt's book (Refs. 22 and 23), these comprised a hosepipe about an inch in diameter, with a mouthpiece like a gas-mask which was pressed to the speaker's face. I use the word "speaker", because I would imagine such a mouthpiece would inhibit breathing for a singer (it covered the nose as well), and it would eliminate "chest tones." It is said that Berliner himself was the vocalist on the first disc records.

Acoustically, the effect would be that there was no bass-cut due to the mouth of the horn, and certainly early Berliners have a "warmth of tone" which seems impossible to explain any other way. However the speaking-tube had its own harmonic resonances, very analogous to those in a conical horn. I did a rough-and-ready experiment with a similar tube coupled to a soundbox, and found it just as easy to detect when a resonance was taking place, and to modify my performance accordingly. I note these details because, if someone wishes to recover the sound from these early records with "fidelity," a different approach may be required.

It is also apparent that some phonographs had a hybrid between a speaking-tube mouthpiece and a proper horn. For example, a small horn only about six inches long might be coupled to a rubber tube. The performance of such an apparatus would depend sharply upon how "soundproof" was the space between the mini-horn and the speaker's face. There would be much more bass-cut if the seal wasn't perfect.

Another variation includes horns not of "conical" shape. (I defined this expression in section 12.3). To judge from photographic evidence, there were comparatively few of these. Most had a curve along their axis, instead of straight boundaries to form a cone. This gave them a shape between that of a trombone and that of an exponential horn. I am very grateful to Adrian Tuddenham for lending me an exponential loudspeaker horn of circular cross-section, which I coupled to a recording-soundbox to study how the non-conical shape would have affected the recording process. When used for recording, the air resonances were at similar frequencies to those in a conical horn of the same effective length and mouth-diameter, but the second and higher harmonics were rather more pronounced. I am afraid my technical knowledge is helpless to diagnose the reason for this - exponential horns are always better for reproduction purposes - but one possibility is a failure in "reciprocity" (section 12.7), because the sound waves are curved in the opposite sense. Lord Rayleigh's definition specifically excludes transitory effects; curvature of sound waves would be significant in this context.

Roland Gelatt's book shows a photograph of a third recording session with another type of horn whose cross-section comprised *two* straight-sided sections. A basic conical horn had an added flare which was also conical (Ref. 24). This picture was taken at Columbia's London studio in 1916, with Alfred Lester, Violet Lorraine, and George Robey posing for the camera. These three artists only made one record together, so we can unambiguously identify the session as the one which produced "Another Little Drink" (Columbia L1034 or 9003). It is rather difficult to estimate the size of the horn (one has to guess the size of George Robey's head), and the neck of the horn is beyond the left-hand edge of the picture; but I reckon it was about 915mm long (965mm when you allow for the end-correction).

Now I am also very grateful to Alistair Murray for lending me a superb copy of the record actually used in Spencer's experiments (Ref. 25). It was recorded for Phoenix in 1911 and later reissued on Regal. It had been chosen for Spencer's research because it was more "tinny-sounding" than most. I decided to compare the two records. Initially they were quite different when played at 80rpm; but when I increased the speed of the

Phoenix to raise the pitch by about two-thirds of a semitone (when, as it happened, the two songs came into the same key), the match seemed very good indeed. The fundamental frequency shown by my hornyness-cure was almost the same as that established by Spencer, and matched a column of air 965mm long. Thus I believe we may have a photograph of the actual horn analysed by his computer. The fact that it was effectively two sections may also explain the uneven spacing of the resonances shown in Spencer's Table 6.2.

This "conical-flared horn" from the "English Columbia" stable features so often on British acoustic records, that we have provided a special switch-setting for it. Because it has a flare, the longitudinal resonances are not only better-distributed, but more damped (because they match the open air better). Similar effects are often audible on continental recordings from the Odeon/Parlophon stable.

A great many horns were conical but with a kink in their central axis to allow the horn to "look at" something. There are four such horns at Hayes, all numbered with the prefix "0". We have already mentioned the "Type 01" used for grand pianos. The "Type 04" appears to be made for duettists, since it comprises two such horns side-by-side in "Siamese Twin" fashion (Ref. 26). Unfortunately the duettists theory must be discarded, because it simply isn't possible for two human beings to get their mouths sufficiently close together - the centres of the two open ends are only nine inches apart! I did an extensive search of the HMV Artists' Sheets to discover what Horn 04 was actually used for, without success. Adrian Tuddenham has suggested that it was designed to look down at a concertina or melodeon, instruments featured on Zonophone rather than HMV.

I am very grateful to John Gomer for pointing out that many hill-and-dale soundboxes (both for recording and reproduction) had a mechanism to allow for variations in the level of the wax. Instead of the pivot of the stylus-bar effectively being fixed, it was free to float up and down. (The same principle may be seen on "fishtail weight" reproducers for cylinders). Vibrations were transferred only because the pivot was on a comparatively massive sub-chassis, forming a "high-pass filter" which cut the bass (namely, the unevenness of the wax). It would be effective below about 100Hz, taking low frequency sounds away from the cutter. So we have a further bass-cut at 12 decibels per octave on the recording in addition to the other bass-cuts. Precisely the same effect would also occur whenever a master-cylinder was copied using a so-called "pantograph". In the latter case, the cut-off frequency can sometimes be estimated from the spectrum of the rumbling-noises transferred from the master; but the cumulative effect is so great that I cannot believe we will ever neutralise it.

However, the French Pathé company (the principal advocates of hill-and-dale discs in Europe) extended similar techniques to their disc records. Master-recordings were generally made on large wax cylinders some five inches (13cm) in diameter, and then transferred to published moulded cylinders and pressed discs as the demand arose. The discs might be of several sizes, including 20cm, 27.5cm, 35cm and 50cm; but not all recordings were issued in all sizes. However, a new possibility exists for reducing the background rumble, because different-sized discs of the same performance would now have different recorded rumble characteristics. Some experimental work on these lines has shown that a small amount of "mechanical amplification" was also introduced for larger discs (35cm and 50cm). Presumably this was achieved by lengthening the cantilever of the cutting-stylus so it functioned as a magnifying lever; so these discs should also have lower surface-noise (all other things being equal).

My final "variation" consists of "pneumatic amplification." About one percent of all lateral-cut commercial acoustic recordings apparently comprise discs which have been

copied through "an Auxetophone" (or equivalent). This was a reproducing machine whose soundbox was replaced by a valve mechanism fed by compressed air; it is reliably reported that such a machine could be heard six miles away. A wax disc would be processed into a low-noise metal "mother", and then played on a studio Auxetophone into another acoustic recording machine. With careful choice of the three horns involved, the resulting disc could sound almost identical to the original, except louder; the technique was particularly used on the continent of Europe for string orchestras. Three things might confirm this has happened. First, the normal bass compensation for the mouth of a recording horn becomes impotent. Second, modern methods of surface noise reduction may reveal a second layer of surface noise behind the first. Thirdly, "peak overloading" may be audible, because the acoustic power at the neck of the Auxetophone drove the air into non-linearity, causing intermodulation products at mid-frequencies.

12.26 When we should apply these lessons

The "harmonic resonances" of the air in a horn occurred at a fundamental frequency and its harmonics; we saw earlier how singers might adapt their techniques. There is no doubt that certain elocutionists in acoustic recording days adapted very successfully - people like Harry E. Humphrey and Russell Hunting, for example. But other speakers obviously didn't, and it may be ethically justifiable to reverse the effect.

The question then arises, who else might have been affected? We mentioned such performance compromises in section 12.4; but I found that my wife, a viola player, was unable to hear the effect when a Type 11 1/2 horn was picking up her instrument. The matter would have been even more difficult with the correct Type 11 1/2A, designed to "look down on" the bridge of a stringed instrument rather than face the musician square-on. The effect was also inaudible to a pianist when a Type 01 horn was placed anywhere near the hammers of a grand piano, although the recording expert may have positioned it away from critical notes of course.

The acoustic process has always appeared more successful with vocal music than other subject matter. This has usually been attributed to the limited frequency range; but now we have another reason. The lesson is that new playback techniques will probably be more successful on instrumental records than vocal ones.

Most other problems faced by artists in the acoustic era also applied to electrical recordings, so I shall leave those considerations until my final chapter.

12.27 Summary of present-day equalisation possibilities

Because we need resonant circuits to compensate for the acoustic recording machinery, we must apply de-clicking techniques before the compensation circuitry. The slightest "grittiness" means the cure will be worse than the disease, because the circuitry will "ring".

At present (as with electrical records), I prefer to convert back to analogue for equalisation purposes, so the relative phases always come right. But I can see that various statistical techniques might be used in the digital domain to quantify resonant frequencies, and give a "figure of confidence" for the results. For multi-horn recordings it might even be possible to do several statistical analyses at once. It remains to be seen how much data

is needed to achieve consistent results - it might be as little as one second or as much as several minutes. But if it is the former, we could then quantify not only how the different horns were "mixed", but when the balance "changed."

Correcting the bass-cut due to the mouth of a horn has sometimes proved difficult, because low-frequency background noise becomes amplified. (to be written) (incl. coloured background noise)

So far I haven't referred to Stockham's method for curing even longer-term resonances, such as those within the metal of the horn (Ref. 8). Stockham uses an empirical technique to distinguish between the short-term equalisation and the long term equalisation. Since we can deal with the (comparatively) short-term effects using conventional techniques, we can keep Stockham's method up our sleeves (it is rather controversial anyway), until we have run out of conventional techniques. Then Stockham's method will not be used as a cure-all, and will get a flying start on the long term equalisation. Alternatively, his method might be reserved for when conventional processes give ambiguous results.

12.28 Conclusion

This chapter has only revealed "the tip of the iceberg." There is a great deal more work to be done. However, the achievements of Spencer and Stockham (shown in sections 12.10 and 12.12) show that it is possible to harness computers to the task of analysing historic sound recordings, for the purpose of discovering the characteristics of the equipment which made them. It isn't the fault of either worker that they could see only part of the picture, so they tackled only part of the problem. And I'm sure I haven't thought of everything either!

Before long, computers programmed to look for the known defects should allow acoustic recordings to sound as if they had been recorded through a microphone. But due to the compromises made by the performers, I am also certain human beings will be needed to override the results when they are artistically inappropriate.

TECHNICAL APPENDIX - now out of date, to be replaced

Prototype equalisers have been built to neutralise the effects of the air in a conical recording horn. A fundamental resonance and seven harmonics were equalised in one circuit, and the bass-cut due to the mouth of the horn in another. When tested against several conical horns under anechoic conditions, correct axial response could be restored within the limits of experimental error (about 1 decibel, and plus or minus one percent in frequency). I was not surprised by this, since Spencer and Olson had separately shown this would be the case (Refs. 4 and 6); but the following extra information came to light

Spencer's expression for relating the fundamental acoustic resonance to the length of the horn was found defective in two respects. First, he did not use "end corrections". I found the usual "$8R/3\pi$" end-correction to be appropriate. Secondly, he did not use a suitable value for the velocity of sound in air; I was obliged to measure the ambient temperature of my experiments before I could verify his formula. When this was taken into account, the fundamental frequencies for a number of horns matched the physical measurements within about 0.5%. Assuming acoustic recording studios were quite warm, I used a value of 345 meters/sec in my calculations for historic sessions.

For the fundamental resonances, the prototype equaliser covered the range 150Hz - 300Hz at half-semitone intervals. This range was about right for most HMV records, although the very long horn with the 114Hz fundamental would not have been covered. The intervals were arranged logarithmically; but in retrospect I think it would have been better to have had them linearly-spaced, because the Q-factors were greater at low frequencies, needing more accurate tuning.

Another reason for the prototype was to discover whether I should compensate for even higher harmonics. I expected these would cause variations of only a couple of decibels or so, probably drowned by other sources of error. But a fundamental and only seven harmonics proved to be insufficient for two reasons. Firstly, higher harmonic resonances were within the range of frequencies which were both recorded most efficiently, and which were near the maximum sensitivity of the human ear. Secondly, the presence of uncorrected higher harmonics could easily be mistaken for those of another horn or of the coupling gadget. The fundamental and perhaps eleven harmonics might have been preferable.

However, it proved perfectly acceptable for the bass-cut due to the mouth of the horn to be adjusted on the same control. Its error was never more than one decibel once the resonances had been compensated, and it was much more convenient to have the two processes working together.

REFERENCES

1: George Brock-Nannestad, "A comment on the instalments by Peter Copeland on Acoustic Recordings", Sheffield: Historic Record (*magazine*), No. 34 (January 1995), pp. 23-24.

2: The "Reciprocity Theorem" was first proved mathematically by Helmholtz, and later extended by Lord Rayleigh. Rayleigh's own verbal exposition is thus: "If in a space filled with air which is partly bounded by finitely extended fixed bodies and is partly unbounded, sound waves be excited at any point A, the resulting velocity-potential at a second point B is the same both in magnitude and phase, as it would have been at A, had B been the source of sound." (J. W. S. Rayleigh, "Theory of Sound," (book, second edition (1896), Article 294).

3: Irving B. Crandall Ph.D., "Theory of Vibrating Systems and Sound" (*book*), Van Nostrand & Co. (New York) or Macmillan & Co. (London), July 1926, pp. 154 - 157.

4: H. F. Olson, RCA Review, Vol. 1, No. 4, p. 68. (1937). The formula is also quoted in Olson's book, which is more easily accessible: Harry F. Olson and Frank Massa, "Applied Acoustics" (2nd edition), P. Blakiston's Son & Co. (Philadelphia) or Constable & Co. (London), 1939, page 221.

5: George Brock-Nannestad, "Horn Resonances in the Acoustico-Mechanical Recording Process and the measurement and elimination in the replay situation," Phonographic Bulletin No. 38, pp. 39-43 (March 1984).

6: Paul S. Spencer, "System Identification with application to the restoration of archived gramophone recordings" (*doctoral thesis*), University of Cambridge Department of Engineering, June 1990.

7: G. A. Briggs, "Sound Reproduction" (book, 3rd edition, 1953), Wharfedale Wireless Works, pp. 267-8.

8: Thomas G. Stockham, Jr; Thomas M. Cannon; and Robert B. Ingebretsen: "Blind Deconvolution Through Digital Signal Processing," Proceedings of the I.E.E., Vol. 63 No. 4 (April 1975), pp. 678-692.

9: (in Great Britain) RCA Red Seal RL11749 (L.P.), RK11749 (cassette).

10: RCA Victor Gold Seal 60495-2-RG. (Boxed set of 12 CDs).

11: Gramophone Company matrix number 4195F. Published pressed in vinyl from original matrix on Historic Masters HMB36. Also dubbed to L.P., Opus 84.

12: George Brock-Nannestad, Presentation at seminar "Audio Preservation Transfer Technology for the Sound Archivist", Peabody Conservatory of Music, Baltimore, 26th June 1991.

13: Paul Whiteman, "Records for the Millions" (*book*), New York, Hermitage Press Inc., 1948, page 3.

14: George Brock-Nannestad, "The Victor Talking Machine Co. new process of recording" *article*), Sheffield: "Historic Record" No. 35 (April 1995), page 29.

15: Percy Wilson M.A. and G. W. Webb, "Modern Gramophones and Electrical Reproducers" (*book*), London, Cassell and Company, 1929, pages 29-32.

16: Wente and Thuras, Journal of the Acoustic Society of America, Vol. III No. 1, page 44. The following two books may be more accessible: Harry F. Olson and Frank Massa, "Applied Acoustics" (2nd edition, 1939), Constable and company, London, pages 103-105; and A. E. Robertson, "Microphones" (2nd edition, 1963), Iliffe Books, London, pages 80-83. The latter describes the British equivalent microphone, known as the "Standard Telephones & Cables Type 4017."

17: Here are some examples. (1) Beka 40628: "The Vision of Salome Waltz" (recorded 1908). Top D from the piccolo, doubling other instruments, is missing. This corresponds to 2343Hz. (2) Columbia L1067 (matrix 6766), "Till Eulenspiegel" (NQHO/Wood), top E flat on first violin disappears, corresponding to 2488Hz. (3) Two HMV records, DB.788 (matrix Cc3683-7 recorded 23-10-23) and DB.815 (matrix Cc5396-2, recorded 1-12-24) both eliminate top F-sharps from different violinists - Thibaud and Kreisler - corresponding to a frequency of 2953Hz. All these frequencies assume the artists were tuned to modern concert-pitch, which is the biggest potential source of error; but the frequencies I have given are certainly correct within three percent.

18: J. W. S. Rayleigh, The Theory of Sound, Volume 1 (second edition of 1894). Chapter X deals entirely with the vibrations of plates, with sections 218-221a specialising in disc-shaped plates. Experimental results are dealt with in section 220.

19: For the HMV pattern, see Peter Ford, "History of Sound Recording," Recorded Sound No. 7, p. 228. (This article says "The recorder box on the Gramophone Company's machine had a cord and coiled-spring tensioning-device that . . . provided a means of tuning the resonances of the diaphragm assembly to fine limits." This is not true. In the late 1980s the cord on the surviving machine perished and the recorder-box fell apart, and for the first time it was possible to see how it worked. In fact, the mechanism served merely to allow the stylus-lever and (more important) the cutter to be changed quickly). Brock-Nannestad has also published pictures of the HMV pattern in two places: a photograph in Gramophone magazine, Volume 61 No. 728 (January 1984), page 927; and sketches of two Victor soundboxes (IASA Phonographic Bulletin No. 38, March 1984.)

20: For the Columbia pattern, see Percy Wilson M.A. and G. W. Webb, Modern Gramophones and Electrical Reproducers (Cassell, 1929), Plate II. Additional information is given in The Gramophone, Vol. 5 No. 12 (May 1928), page 488.

21: Two examples: (1) Liszt's Hungarian Rhapsody, performed by the London Symphony Orchestra conducted by Arthur Nikisch, and recorded by HMV at Hayes. The clash occurs between sides 2 and 3, recorded on 21st June 1914 and 25th June 1913 (matrixes Ho 564c and Ho 501c).

(2) Scriabin's "Poème d'Extase", performed by the London Symphony Orchestra conducted by Albert Coates on Columbia L.1380-2. Clashes occur between sides 2 and 3, and between 4 and 5. The work was recorded on 27th April 1920, but sides 3 and 4 were retaken on 7th May 1920.

22: Roland Gelatt, "The Fabulous Phonograph" (2nd edition), Cassell & Co., London, 1977, page 62.

23: Ibid., plate of Sarah Bernhardt following page 64.

24: Ibid., plate of the stars of "The Bing Boys" following page 64.

25: Illustrated London News, 21st December 1907.

26: Benet Bergonzi, "Gas Shell Bombardment", HR 17 (October 1990), pages 18-21; and Peter Adamson, "The Gas Shell Bombardment Record - some further thoughts," HR 19 (April 1991), pages 13-15.

13 The engineer and the artist

13.1 Introduction

Until now, I have concentrated upon recovering the original sound (or the *intended* original sound) from analogue media. But now we enter what, for me, is the most difficult part of the subject, because we must now study the ways in which the artists modified their performances (or their *intended* performances) for the appropriate media. I find this difficult, because there is little of an objective nature; effectively it is *subjectivism* I must describe.

But I hope you agree these subjective elements are very much part of sound recording history. Since the industry has undergone many revolutions (both technological and artistic), I hope you also agree that understanding these subjective elements is a vital part of understanding what we have inherited. It may also help you understand what we have *not* inherited - why some people were considered "born recording artists" and others weren't, for example. (And I shall continue to use the contemporary word "artists", even though the word "performers" would be better).

The structure of this chapter will basically be chronological, although it will switch halfway from pure sound to sound accompanying pictures. For the period 1877 to 1925 (the approximate years of "acoustic recording"), we examined some of the compromises in the previous chapter. We lack a complete list of the methods by which artists and record producers modified performances for the recording horn(s), but it is clear the idea was to *simulate* reality using all the tools available, including hiring specialist vocalists, elocutionists, and instrumentalists. The effect upon the performers was totally untypical, and we saw that this raises severe problems for us today. Should we be even be *trying* to reproduce "the original sound"?

We also saw how the acoustic process revolutionised popular music by dictating the three-minute dance hit, its instrumentation, and its simple rhythmic structures. On the other hand, Edmundo Ros' complex Latin-American rhythms were almost inaudible before full frequency range recording was developed.

Meanwhile in classical music, acoustic repertoire was biassed by the lack of amplification. This prevented any representation of large choral works, any use of instruments dating from the eighteenth century or before, or the employment of massed strings or other large sound sources such as organs. Whilst a cut-down symphony orchestra with (perhaps) eight string instruments was certainly feasible in the 1920s, it still remained common for brass bands and special recording ensembles to perform repertoire composed for a classic symphony orchestra.

But with electronic amplification, for the first time the performers could "do their own thing" without having to modify their performances quite so much. And the *art* of sound recording could begin.

13.2 The effect of playing-time upon recorded performances

By 1902, Berliner's seven-inch disc records had become outgrown, although several companies continued to make seven-inch discs for the niche market of seven-inch

gramophone turntables. The two dominant sizes for the next fifty years were ten-inch and twelve-inch records. (I shall continue to use these expressions rather than the metric equivalents, because (*a*) they were notional sizes only, some records being slightly larger and some slightly smaller; and (*b*) when official standards were established in 1954-5, the diameters weren't "round figures", either in inches or millimetres!)

Seven-inch discs and two-minute cylinders continued until 1907; but the "standard song" (inasmuch as there is such a thing) might be cut to one verse and two choruses only - a mere travesty of the ten-inch version, which could have a much better structure. The three-minute melody became the cornerstone of the popular music industry for the next seventy-five years. It survived the transition to seven-inch 45rpm "singles," and was only liberated by twelve-inch 45rpm "disco singles" in 1977.

Three-minute melodies were also optimum for dancing. These always had a distinct ending with a major chord of the appropriate key, signalling dancers to stop and applaud each other. From about 1960 many songwriters had come to hate this type of ending, and introduced "the fade-ending" (done electronically) to eliminate the burden. But radio disc jockeys, desiring to "keep the music moving" while promoting their show, began to "talk over" the fade ending as soon as its volume allowed.

At this point I shall introduce a story about the "protest singer" Bob Dylan. In 1965 he decided to break the boundaries of the "three-minute single", and asked his company (US Columbia) the maximum duration which could be fitted on a 45rpm side. He was told "six minutes", this being the duration of an "extended play" record (with two tunes on each side). So he wrote *Like A Rolling Stone* with a very unusual structure, so it would be impossible for Columbia to cut the master tape with a razor blade. Each verse and each chorus started one beat ahead of a bar line, so an edit would scramble the words. It had twenty-bar verses, followed by twelve-bar choruses, and twelve-bar instrumental "link passages". Although I have no evidence, I surmise he deliberately blew the "link passages" on his harmonica badly, so no-one could ever edit two link passages together and shorten the song that way! The result ran for precisely five minutes and fifty-nine seconds - a specific case of sound recording practices altering a performance.

Twelve-inch "78s", introduced in 1903, played for between four and five minutes per side. When vocal records predominated, this was adequate; hardly any songs or arias were longer than that. But all these timings are very nominal. Any collector can think of records considerably shorter than these, and Peter Adamson has listed many of the longest-playing 78rpm discs (Ref. 1). Some of the latter illustrate the point I made in section 5.4 about engineers deliberately slowing their recording turntable, to help fit a long tune onto one side.

But I raise this matter so we may understand what happened when pieces of music *longer* than four minutes were recorded. At first the music would be shortened by the musical director brutally cutting it for the recording session. It was not until 20th November 1913 that the first substantially uncut symphony was recorded (Beethoven's Fifth, conducted by Nikisch) (Ref. 2). This established the convention that side changes should be at a significant point in the music, usually at the end of a theme. But when a chord linked the two themes, it was common practice to play the chord again at the start of the new side. This is a case where we might have to make significant differences between our archive copy and our service copy.

I shall now point out a couple of less obvious features of this particular situation. In Britain, that symphony was first marketed on eight single-sided records priced at six shillings each. If we rely on official figures for the cost of living (based on "the necessities of life"), this translates in the year 2000 as almost exactly one hundred and fifty British

pounds. Shortly afterwards it was repackaged on four double-sided records at six shillings and sixpence each, so it dropped to about eighty-two pounds. Although it isn't unique in this respect, Beethoven's "Fifth" could be divided into four movements, each of which occupied one double-sided record. Therefore it was feasible for customers to buy it a movement at a time - for slightly over twenty pounds per movement!

Although I described it as "substantially uncut", the exposition repeats of both the first and last movements have gone. In 1926 Percy Scholes wrote: "In the early days of . . . 'Sonata' Form it was usual to repeat the Enunciation. This was little more than a convention, to help listeners to grasp the material of the Movement" (Ref. 3). Because both repeats would have added ten pounds to the price, and pedants could always put the soundbox back to the start of the side at the end of the first exposition, you can see that quite a number of issues arose from the cost, the limited playing time, and how the music was divided up.

I apologise if this section has concentrated upon Western "classical" music, but that is what I find myself handling most frequently. Of course, many other types of music can be affected. For example, the modifications to performances of Hindustani classical music (in this case, very severe ones) have recently been studied by Suman Ghosh. (Ref. 4)

In more complex cases, a service copy might tamper with the original performance very significantly. One such case is where there was "an overlap changeover." This writer was once involved in acquiring a performance of Bach's Brandenburg Concerto No.2, during which an outgoing theme overlapped an incoming theme. The work had been conducted by a significant British conductor; but the recording had remained unpublished, and unique test pressings were made available by the conductor's widow. She was quite adamant that the repeated bar should be axed; and this was before the days of digital editors which make such a task simple! So some nifty work with a twin-track analogue tape recorder was necessary to get the music to sound continuous. Readers will have to decide the solutions to such problems themselves; but this was the actual case in which "cultural pressures were so intense" that I could not do an archive copy, as I mentioned in section 2.5.

The other complex case is where disc sides changed, forcing a break in the reproduction. Some conductors minimised the shock of a side change by performing a ritardo at the end of the first side. (Sir Henry Wood was the principal proponent of this technique). Of course it makes complete musical nonsense with a straight splice; but we now have technology which allows a change of *tempo* to be tidied, thereby presumably realising the conductor's original intentions. Again, you have to understand why these performance modifications occurred before you can legitimately recreate "the original *intended* sound."

Yet record review magazines definitely show that *customers* weren't happy with music being broken up like this. To reduce their concerns, commercial record companies introduced the "auto coupling" principle. Large sets of records could be purchased in an alternative form, having consecutive sides on different physical discs. An "auto changer" held a pile of discs above the turntable on an extended spindle, and automatic mechanisms removed the pickup at the end of each side, dropped another disc on top of the first, and put the pickup into the new groove. When half the sides had been played, the customer picked up the stack, turned it over without re-sorting the records, and loaded it on the spindle for the second half of the music.

If your institution has such records, you should remember that the "manual" version and the "auto coupling" version can give you two copies of the same recording,

usually under different catalogue numbers. (This may help with the power-bandwidth product). As a small aside to this issue, I personally maintain that when 78rpm sides were sold in auto coupling versions, this is definite evidence that the conductor (or someone, at least) *intended* us to join up the sides today.

Most long music continued to be recorded in "four-minute chunks" until about 1950. If the subject matter was intended to be a *broadcast* rather than a disc record, obviously we must rejoin the sides to restore "the original intended sound," even though it might have been split into four-minute chunks for recording purposes. Nothing else was possible when the power-bandwidth product of 78s meant sounds could only be copied with an audible quality loss. A breakthrough of sorts occurred in about 1940, when American Columbia started mastering complete symphonic movements on outsized (44cm) nitrate discs, probably using the same technology as "broadcast transcriptions" used in the radio industry. This seems to have been in anticipation of the LP (Long Playing record). I say "breakthrough of sorts", because although the nitrate had a better power-bandwidth product (in the power dimension, anyway), it meant a note perfect performance for fifteen minutes or more. Practical editing still wasn't feasible.

Copy-editing of disc media seems to have been pioneered by the Gramophone Company of Great Britain in 1928, but the results were not very convincing for quality, and almost every company avoided it whenever possible. "Auto couplings" reduced the effects of side changes, but still meant one large break in the music where the stack of records had to be turned over. For broadcasters this was intolerable; so they made disc recordings with their sides arranged in a different pattern, known as "broadcast couplings." They assumed *two* turntables would be available for transmitting the discs, and the odd-numbered sides would come from the discs on one turntable and the even-numbered sides from the discs on the other, while consecutive sides were never on the same physical disc. These couplings resulted from continuous disc recordings made with two disc-cutting turntables; and if the recording was made on double-sided blanks, "broadcast couplings" automatically resulted. This writer regards these situations as meaning we *must* join up the sides on the service copy.

Broadcasters often used "overlap changeovers" - passages recorded in duplicate on the "outgoing" and "incoming" discs. The incoming disc (heard on headphones) would be varispeeded manually during a broadcast, until it synchronised with the "outgoing" disc; after which a crossfade from one to the other would be performed. Today we simply choose the passage with the better power-bandwidth product, and discard the other version.

Meanwhile, the craft of disc editing was perfected by broadcasters. "Jump cuts" were practised - the operator would simply jump the stylus from one part of a disc to another, in order to drop a question from an interview for example. Surviving nitrates often bear yellow "chinagraph" marks and written instructions showing how they were meant to be used. And the disc medium permitted another application for broadcasters, which vanished when tape was introduced - the "delayed live recording." It was quite practicable to record (say) a football commentary continuously. When a goal was scored, the pickup could be placed on the disc (say) thirty seconds earlier, and the recording was reproduced while it was still being made. (Only now is digital technology allowing this to happen again).

Optical film, and later magnetic tape, allowed "cut and splice" editing. I shall talk about the issues for film in section 13.13, but it was quickly realised by English Decca that you couldn't just stick bits of tape together to make a continuous performance for long-playing records. If the music just stopped and started again (which happened when

mastering on 78s), a straight splice meant the studio reverberation would also suddenly stop, very disturbing because this cannot occur in nature. When mastering on tape, it was vital to record a bar or two of music preceding the edit point, so the studio would be filled with suitable reverberation and a straight splice would be acceptable.

From this point, reviewers and other parties started criticising the process of mastering on tape, because record companies could synthesise a perfection which many artists could not possibly give. However, this criticism has evaporated today. The public *expects* a note perfect performance, and the art of the record producer has evolved to preserve the "spirit" of an artistic interpretation among the hassles of making it note perfect. Thus, such commercial sound recordings and recorded broadcasts differ from reality, and this branch of music is now accepted as being a different medium.

13.3 The introduction of the microphone

Apart from telephony (where high fidelity was not an issue), microphones were first used for radio broadcasting in the early 1920s. This was the first significant use of electronic amplification. Although early microphones lacked rigorous engineering design, it is clear that they were intended to be channels of "high fidelity" (although that phrase did not come into use until about 1931). That is to say, the microphone was meant to be "transparent" - not imposing its identity on the broadcast programme - and microphones rapidly evolved to meet this criterion.

Customers soon recognised that the commercial record industry had better musicians than broadcasters could afford. On the other hand, broadcasts lacked the constant hiss and crackle of records, and any "megaphone quality" added to artists. Record sales began to decline against the new competition.

As we saw in chapter 5, the solution was developed by Western Electric in America. Using rigorous mathematical procedures (itself a novel idea), they designed apparatus with rigorous behaviour. Much of this apparatus was used for public address purposes at the opening of the British Empire Exhibition in 1924, which was also the first time the voice of King George V was broadcast; this survives today in a very poor acoustic recording! Electrical methods of putting the sound faithfully onto disc had to wait a little longer, until the patent licensing situation was sorted out. As far as I can establish, the earliest Western Electric recording to be published at the time was mastered by Columbia in New York on 14th February 1925 (Ref. 5). Nobody (myself included) seems to have noticed the significance of its recorded content. It comprised a session by "Art Gillham, the Whispering Pianist"! Although he clearly isn't whispering, and the voice might have been captured by a speaking tube instead of an acoustic recording horn with the same effect as a microphone, it would have been impossible to publish his accompanying himself on the piano any other way.

A performance technique which was often believed to depend on technology is "crooning." Crooners began in radio broadcasting, when listeners were wearing headphones to avoid their needing electronic amplification; this helped the implication of a vocalist singing quietly into a listener's ear. But the art of the "crooner" soon became completely inverted. Before the days of public address, successful crooners had to be capable of making themselves heard like operatic tenors, while still projecting an intimate style of vocal delivery. Contemporary writers frequently criticised crooners for being vocally inept, comparing them with opera singers where "Louder" was often considered "Better"; but in fact, a crooner's craft was very much more complex and self-effacing.

For some years, microphones suffered because they were very large, which affected their performance on sounds arriving from different directions; I shall discuss the significance of this in the next section. As I write this, another consideration is that microphones cannot be subject to unlimited amplification - most microphones have less sensitivity than a healthy human ear. I know of only one which can just about *equal* the human ear. It is made for laboratory noise measurement applications, but it is a big design, so it suffers the size disadvantage I've just mentioned. In practice, the best current studio microphones have a noise performance a full 8dB worse than the human hearing threshold. So, for high fidelity recording which also takes into account "scale distortion", we still have a little way to go.

13.4 The performing environment

Since prehistoric times, any sounds meant to have an artistic impact (whatever that means!) have evolved to be performed in a space with certain acoustic properties. This is true whether the location is a cathedral, an outdoor theatre, or a woodland in spring. It is an important factor which may influence the performances we hear today. It took nearly half a century for sound recording technology to be capable of handling it. From 1877 until 31st March 1925 nearly all sound recordings were made without any worthwhile representation of the acoustic environment. On that date, US Columbia recorded six numbers from a concert given by the Associated Glee Clubs of America in the Metropolitan Opera House, New York. (Ref. 6).

Although these had been preceded by radio broadcasts from outside locations, they were the first published recordings to demonstrate that the acoustic environment was an important part of the music. And another decade went by before even this was widely appreciated. It wasn't just that the end product sounded better. Musicians and speakers were also affected, and usually gave their best in the correct location. I shall now give a brief history of the subject. In fact, it is a very complex matter, worthy of a book on its own; I shall just outline the features which make it important to restoration operators.

I can think of only one acoustic recording where the surrounding environment played a significant part - the recording of the Lord Mayor of London, Sir Charles Wakefield, in 1916 (Ref. 7). In this recording it is just possible to hear a long reverberation time surrounding the speech, which must have been overwhelming on location. Apparently the Gramophone Company's experts thought this was so abnormal that they took the precaution of insisting that "Recorded in Mansion House" was printed on the label. As far as I can ascertain, this was the first case where the location of a commercial recording was publicly given, although we now know many earlier records were made away from formal studios.

When the microphone became available, together with the electronic amplifier, it was possible to get away from the unnatural environment of the acoustic recording studio. Because there was a continuously rolling recording programme, it took several months for the new advantages to be realised. In the meantime there were some horrible compromises. Serious orchestral performances, made in acoustic recording studios and recorded electrically, sounded what they were - brash, boxy, and aggressive. It was no longer necessary for artists to project their performances with great vigour to get their music through the apparatus, but they continued to perform as though it was. Sir Compton Mackenzie's comments about the first electrically recorded symphony have

been much quoted (Ref. 8), but anyone hearing the original recording (even when correctly equalised) must agree with him.

It was first necessary to get away from the acoustic recording rooms. In Britain some choral recordings were tried from the "Girls' Canteen" at the factory in Hayes. Next, recording machinery was installed at Gloucester House in the centre of London, and various locations were tried by means of landlines. Soon, half a dozen halls and places of worship in London became recording studios on a temporary or permanent basis. Before the end of the year, some unsung genius discovered The Kingsway Hall, which became London's premier audio recording location until the late 1980s. The first half of 1926 was the heyday of "live recording," with performances from the Royal Albert Hall and the Royal Opera House being immortalised; but the difficulties of cueing the recording, the unpredictability of the dynamics, and the risk of background noises, slowed this activity. Instead, the bias turned in favour of recordings made from natural locations to the machine in *private* performances, not public ones.

Early microphones were essentially omnidirectional, so they picked up more reverberation than we would consider normal today. Furthermore, their physical bulk meant they picked up an excess of high frequencies at the front and a deficient amount at the sides and back. This affected both the halls which were chosen, and the positioning of the mikes. Since we cannot correct this fault with present-day technology, we must remember that "woolly bass" is quite normal, and contemporary artists may have modified their performances to allow for this.

Volume changes in a performance become evened out when more reverberation is picked up. This reduces transient overloads (particularly troublesome on grooved media). All this explains why we may find recordings which are both "deader" and "livelier" than we would normally expect today.

The presence of any reverberation at all must have seemed very unusual after acoustic recording days. This explains what happened after the world's first purpose-built recording studios were opened at Abbey Road in 1931. The biggest studio, Studio 1, was intended for orchestras and other large ensembles; and the first recorded works, including "Falstaff" and the Bruch Violin Concerto, show that its sound was the equal of present-day standards. But then some "acoustic experts" were brought in, and the natural reverberation was dampened. The following year similar recordings were done in the same place with a reverberation time about half its previous duration; and this remained the norm until about 1944. Many artists hated the acoustics so much that sessions had to resume at the Kingsway Hall.

The two smaller studios suffered less alteration. Studio 2 (used for dance bands) seems to have been deliberately designed to work with the microphones, since it had a surprisingly even amount of reverberation at all frequencies. It was longer than present-day fashion, but is not grossly distasteful to us. Studio 3 was intended for solo piano, speech, and small ensembles up to the size of a string quartet. It was practically dead, and you can hardly hear the acoustics at all. It may well have affected the performers, but there is nothing to show it.

Microphones with frequency responses which were equally balanced from all directions came into use in 1932 - the first "ribbon" microphones, with a bi-directional polar diagram. This was not intuitive, and many artists found it difficult to work with them. An omnidirectional equivalent with similarly uniform directionality did not appear until 1936 (the "apple and biscuit" microphone, or Western Electric 607).

Probably the creative use of acoustics reached its peak in radio drama. In radio there are only three ways to indicate the "scenery" - narration, sound effects, and

acoustics. Too much of the first two would alienate listeners, but acoustics could provide "subliminal scenery." When Broadcasting House opened in London in 1932, a suite of drama studios with differing acoustics could be combined at a "dramatic control panel". There was also a studio for live mood music, and another for the messier or noisier sound effects. An entertaining account of all this may be found in the whodunit *Death At Broadcasting House* (Ref. 9), but the art of deliberate selection of acoustics did not cease in 1932. The arrival of ribbon microphones meant that actors could also give "depth" to their performances by walking round to the side and pitching their voices appropriately; even in the 1960s, this was a source of considerable confusion to newcomers, and scenes had to be "blocked" with the rigour of stage performances. Using this technique, the listener could be led to identify himself with a particular character subliminally. Meanwhile, studio managers were manipulating "hard" and "soft" screens, curtains, carpets, and artificial reverberation systems of one type or another, to get a wider variety of acoustics in a reasonable amount of space (Ref. 10). It is the writer's regret that these elements of audio craftsmanship never took root in America.

An "echo chamber" formed one of the radio drama rooms at Broadcasting House. It contained a loudspeaker and a microphone in a bare room for adding a long reverberation time (for ghosts, dungeons, etc). It was not particularly "natural" for music purposes, but film companies soon adopted the idea for "beefing up" the voices of visually attractive but less talented singers in musical films. The Abbey Road studios had an echo chamber of their own for popular vocalists from the mid-1940s, and so did other recording centres where popular music was created. But, to be any good, an echo chamber had to be fairly big; and it was always liable to interference from passing aircraft, underground trains, toilets being flushed, etc. Engineers sought other ways of achieving a similar effect.

One of the most significant was a device introduced by the Hammond Organ Company of America - the "echo spring." This basically comprised a disc cutting-head mechanism generating torsional vibrations into a coil of spring steel wire, with a pickup at the other end. Hammond made several sizes of springs with different reverberation times. They were used in early electronic musical instruments, and later by enterprising recording studios, and the only difficulty was that they tended to make unmusical twanging noises on transient sounds.

This difficulty was eliminated by the "reverberation plate", invented by Otto Kühl in 1956 (Ref. 11). This comprised a sheet of plain steel, two metres by one and a millimetre thick, suspended in a frame, with a driver in the middle and a pickup on one edge. An acoustic damping plate nearby could be wound closer or further away to mop up sounds emitted by the steel, and caused the reverberation time to vary from about 1 second to over 4.5 seconds at will. While it was not completely immune from interference, it was much less susceptible. And because it was smaller than a chamber and could do a wider range of acoustics, nearly every studio centre had one.

In the early 1970s digital signal processing technology came into the scene. The "Lexicon" was probably the first, and was immediately taken by pop studios because it permitted some reverberation effects impossible in real life. Today, such units are made by many companies with many target customers at many prices, and developments are still happening as I write.

There are two points I want to make about all these systems - the "chamber," the "spring," the "plate" and the "digital reverb." The first is that they all *add* reverberation to an existing signal, and if the first signal has reverberation of its own, you get two different lots, which is somewhat unnatural. The other is that the same quantity of

reverberation is added to all the components of that signal - so if it is applied to a vocalist with an orchestral backing, *both* get the same amount of reverberation, irrespective of whether one is supposed to be in the foreground and the other in the background. Although it is possible to limit these disadvantages by means of filters, and by means of "echo send buses" on sound mixing consoles, you should know they exist. I mention them so you may be able to detect when someone has interfered with a naturally reverberated signal and modified the "original sound."

Since most artists instinctively pitch their performances according to the space to be filled, there is rarely any doubt whether a performance was being done "live" or especially for the microphone. But it could become blurred in the mid-1930s, when public address systems first became normal in the bigger halls (Ref. 12). Yet there is another stabilising factor - the effect of reverberation time on the *tempo* of music or speech. If a performance goes too fast in a live acoustic, the notes or words become blurred together, and a public address system cannot always save the situation. The lesson for us today is that if we hear a conventional performing artist in an unusually "live" or an unusually "dead" environment, we cannot assume that the *tempo* of his performance will be "normal."

13.5 "Multitrack" issues

Anyone who asserts that "multitrack recording" was invented in the 1960s is wildly wrong; but I shall be considering its birth (which was for the film industry) in section 13.13. As far as I am aware, the first multitracked pure sound recording was made on 22nd June 1932 by a banjo player called "Patti" (real name Eddie Peabody). On a previous date he had recorded the first layer of sound. This was then processed into a shellac pressing, and played back on a "Panatrope" (an electrical reproducer), while Patti added a second layer. This record (UK Brunswick 1359) was given to me by a musician who was completely unable to understand how Patti had done it!

A month later RCA Victor added a new electrically recorded symphony orchestra to a couple of acoustically recorded sides by Enrico Caruso (section 12.21). In September 1935 the soprano Elisabeth Schumann recorded *both* title parts of the evening prayer from "Hänsel and Gretel" by Humperdinck. In all these cases, the basic rhythm had been laid down first, so the second layer could be performed to fit it.

After the war, magnetic tape recording (which gave instant playback) eliminated the processing delay with disc (we shall cover that problem in sections 13.9 and 13.18). A most remarkable system was invented by guitarist Les Paul, who was joined by his wife Mary Ford as a vocalist (or several vocalists). Les Paul had one of the earliest Ampex professional (mono) tape recorders, and he had its heads rearranged so the playback head was first, then the erase head, and finally the record head. He planned out his arrangements on multi-stave music paper, then normally began by recording the bass line which would suffer least from multiple re-recordings. He and his wife would then sight read two more staves as the tape played back, mixing the new sounds with the original, and adding them to the same piece of tape immediately after it had been erased. By this process they built up some extremely elaborate mixes, which included careful use of a limiter to ensure all the parts of Mary Ford's vocals matched, "double-speed" and "quadruple-speed" effects, and "tape echoes."

Magnetic tape mastering caused the original idea to be reborn in Britain in 1951, using two mono tape machines. This avoided the loss of the entire job whenever there

was a wrong note. For a three-month period EMI Records thrashed the idea, but without any great breakthroughs. Perhaps the most ambitious was when Humphrey Lyttelton emulated a traditional jazz band in "One Man Went To Blow."

Coming to the mid-sixties, multitrack recording was introduced largely because it was necessary to mix musical sources with wide differences in dynamics. So long as the balance between the different instruments could be determined by the teamwork of arranger, conductor, and musician, the recording engineer's input was secondary. But electrically amplified instruments such as guitars meant the natural vocal line might be inaudible. Also many bands could not read music, so there was a lot of improvisation, and accurate manual volume controlling was impossible. Early stage amplifiers for vocals resulted in nasty boxy sounding recordings, so vocalists were persuaded to perform straight into high fidelity microphones in the studio. The engineers re-balanced the music (usually helped by limiters to achieve predictable chorus effects), added reverberation, and supplied a small number of exotic effects of their own making (such as "flanging.") All this only worked if the "leakage" from one track to another was negligible; thus pop studios were, and still are, very well damped. Another consequence was that vocalists were deprived of many of the clues which helped them sing in tune; realtime pitch correctors weren't available until about 1990.

The next British experiments involved the same system as Humphrey Lyttelton a decade before, but using a "twin track" tape recorder, in which either track could be recorded (or both together). The Beatles LP "With The Beatles" was compiled this way in 1963, with the instrumental backing being recorded on one track, and the vocals added on the other. In Britain, the resulting LP was issued in "stereo" (Parlophone PCS3045), which is really a misnomer. There is no stereo "spread"; all the backing is in one loudspeaker, and the vocals in the other!

Within a couple of years, multitrack recording allowed "drop-ins" for patching up defective passages within one track, while international magnetic tape standards allowed musicians on opposite sides of the world to play together. To make this work, the recording head would be used in reverse as a playback head, so the musicians could hear other tracks while everything remained in synchronism. Modern studio equipment is specifically designed to provide very sophisticated sounds on the musicians' headphones. Not only does this enable them to hear their own voices when the backing is excitingly loud, but a great deal of reverberation is sometimes added so they get the "third vital clue" (section 12.4) to help them sing in tune. This may be used even though no reverberation is contemplated for the final mix. Very sophisticated headphone feeds are also needed so each musician can hear the exact tracks he needs to keep exactly in tempo, without any other musicians or a conductor.

By the late 1960s, the creative side of popular music making was using all the benefits of multitrack tape and sophisticated mixing consoles. This enabled musicians to control their own music much more closely, although not necessarily advantageously. The creative input of a record producer became diluted, because each musician liked to hear himself louder than everyone else, with arguments vigorously propounded in inverse proportion to talent! So the record producer's role evolved into a combination of diplomat and hype-generator, while popular music evolved so that musical devices such as melody rubato and dynamics were sacrificed to making the music *loud* (as we saw in section 11.6). The actual balance would be dictated by a "consensus" process. It did not mean sterility; but it slowed the radical changes in popular music of previous decades. This writer sheds an occasional tear for the days when talented engineers like Joe Meek or Geoff Emerick could significantly and *creatively* affect the sound of a band.

In the 1970s, some downmarket pop studios had a poster on the wall saying "Why not rehearse *before* you record?" This was usually a rhetorical question, because all good pop studios were at the frontiers of music. They had to be; it might take several weeks before any results could appear in shops. But soon the conventional idea was soon being stood upon its head. Instead of a recording studio emulating the performance of a group in front of the public, the public performances of musicians emulated what they had achieved in the recording studio. Most of the technology of a recording studio suddenly found itself onstage for live performances. The sophisticated headphone monitoring needed in a dead studio became the "monitor bins" along the front of the stage. These were loudspeakers which were not provided for the benefit of the audience, but merely so the performers could hear each other. Meanwhile, at dance halls, the live band disappeared; and "rap", "scratching" and the heavy unremitting beat brought music to the stage where melody rubato and dynamics were totally redundant.

The technology of "live sound" engineers was soon being applied in the world of classical music - particularly in opera, where the sheer cost of productions meant that sophisticated public address systems meant bigger audiences could be served. Strictly speaking, *recording* engineers did not influence this trend in music. But the recording engineers and the live sound engineers now work together, and between them they will affect the history of music (and drama) even more radically than in the past.

13.6 Signal strengths

In previous chapters we saw that "louder is always better." Today, people doing subjective reviews of hi-fi equipment are well aware that if one system is a fraction of a decibel louder than another, it will seem "better", even though human hearing can only detect a change of a couple of decibels on any one system. As a corollary, acoustic gramophone designers were straining to improve the *loudness* of their products for many years, rather than their frequency ranges.

Yet it was some time before electrically recorded discs were issued at a volume later considered normal. We saw in chapter 5 that there was sometimes a strict "wear test." Test pressings were played on an acoustic gramophone; if they lasted for thirty playings without wear becoming audible, they might be considered for issue; and if they lasted one hundred playings, that "take" would definitely be preferred. The result, of course, was that louder records *failed* the test. So most published records were many decibels quieter than the Western Electric or Blumlein systems could actually do. We see this today, when unissued tests sometimes appear which present-day technology can replay very easily. They sound particularly good, because the dynamics are less restricted, and the signal-to-noise ratio is better. The message to us is that the published version might *not* necessarily be the artist's or producer's preferred choice.

Volumes drifted upwards in the 1940s and 1950s, and limiters helped make records consistently loud as we saw in Chapter 10. The ultimate was reached when recording technology significantly affected the course of musical history in 1963. Quite suddenly, the standard "pop group" of drums, vocals, and three electric guitars (lead, rhythm, and bass), became the norm. Standard musical history attributes this to the success of The Beatles. But this does not explain why similar instrumentation was adopted for genres other than the teenage hearthrob market, such as "surfing" groups, country-and-western bands, rhythm and blues exponents, and even acoustic protest singers like Bob Dylan. The reason, I believe, is the RIAA Equalisation Curve (chapter 5). Yes, I'm

afraid that's right. Seven-inch popular "singles" recorded to this equalisation could hold this instrumentation very precisely without distortion at full volume, so the "louder is better" syndrome was again involved. Indeed, this writer was cutting pop masters in those years, and can swear that the ease of cutting the master was directly proportional to how closely the performance matched this norm.

The subliminal louder-is-better syndrome keeps cropping up throughout the whole of sound recording history, and I will not repeat myself unnecessarily; but I shall be enlarging the topic when we talk about sound accompanying pictures. But more recently, we have become aware that the reproducing equipment of the time also played a part. For example, melody harmony and rubato could be reproduced successfully on 1960s record players of the "Dansette" genre. When more ambitious sound systems became available in the mid-1970s, these musical features became less important. A heavy steady beat became the principal objective; a "Dansette" could never have done justice to such a beat.

In section 13.2 we noted how the maximum playing times of various media influenced what became recorded, and how the three-minute "dance hit" ending with a definite chord led the market from 1902 to the mid-1960s. We also saw how others preferred to escape this scenario by performing an electronic "fade" at the end. In 1967, The Beatles were to demolish the whole idea by publishing a song "Strawberry Fields Forever" with a "false fade ending." Any radio disc jockey who tried introducing the next record over the fade would find himself drowned by the subsequent "fade up", while the fame of the group meant there would be thousands of telephoned complaints at the switchboard. So here is another example of how artists were influenced - creatively - by sound engineering technology.

13.7 Frequency ranges

This brings us to the matter of frequency responses. In principle, engineers could always affect the frequency response of an electrical recording by using circuits in a controllable manner. Yet the deliberate use of equalisation for artistic effect seems to have been a long time in coming. To judge from the historical recorded evidence, the facility was first used for *technical* effect, rather than artistic effect. The situation is reminiscent of 1950s hi-fi demonstrations, comprising thundering basslines and shrieking treble that were never heard in nature. In 1928 the Victor Company made a recording of the "Poet and Peasant" Overture with bass and treble wound fully up (Ref. 13). But the effect was constrained by the limitations of the Western Electric cutter system (which was specifically designed to have a restricted bandwidth), and the effect sounds ludicrous today. I have no evidence for what I am about to say, but it is almost as if Victor wanted a recording which would sound good on a "pre-Orthophonic" Victrola.

According to fictional accounts, controllable equalisation was first used in the film industry in attempts to make "silent" movie stars sound better; but I have no documentary proof of this. Yet watching some early films with modern knowledge shows that relatively subtle equalisation must have been available to match up voices shot at different times and on different locations.

Until the loudspeakers owned by the public had reasonably flat responses, there was little point in doing *subtle* equalisation. Engineers working in radio drama, for example, had to use brutally powerful filtering if they wanted to make a particular point (a telephone voice, or a "thinks" sequence). Conventional equalisation as we know it

today was unknown. Indeed, the researchers planning such installations had only just started to learn how to get a wide flat frequency response. They considered technically illiterate artistic nutters would be quite capable of misusing the facility to the detriment of their employers. This was definitely the case when I entered the industry as late as 1960; both in broadcasting and in commercial recording, the only way we could affect the frequency response was by choice of (and possibly deliberate misuse of) the microphone(s). In any case, the philosophy of today's restoration operator (the objective copy should comprise "the original *intended* sound") means that we hardly ever need to attack such equalisation, even when we know its characteristics.

However, it might be worth noting one point which illuminates both this and the previous section. The louder-is-always-better syndrome resulted in the invention of something caused "the EQ mix." This took advantage of multitrack recorders, and allowed the band to participate in the mixing session in the following way. Equalisation controls on each track could be adjusted to emphasise the most characteristic frequencies of each instrument, for example higher frequencies to increase the clarity of a rhythm guitar. (This would normally be done with something called "a parametric equaliser", which can emphasise a narrow or a wider range of frequencies by a fixed amount. It is actually a wrong word, because it is not the kind of "equaliser" to *equalise* the characteristics of something else, and it does not restore relative phases). If each track had different frequencies emphasised, the resulting mixed sound would be louder and clearer for the same peak signal voltage, and the louder-is-better syndrome was invoked again.

13.8 Monitoring sound recordings

This usually means listening to the sound as the recording (or mixing) is taking place. (It also means "metering" when applied to electrical mixing, but I shall not be thinking about meters in this section).

Professional recording engineers could not monitor acoustic recordings rigorously, although they got as near to it as they could. The musicians were put in a bare room, as I mentioned earlier; but the room would be divided in two by a full height partition. The experts worked on the other side of this with the actual recording machine, and the recording horn poked through a relatively small hole in the wall. Thus the experts heard the music through this hole, and (with experience) they were able to detect gross errors of balance. Communication with the artists was usually by means of a small hinged window or "trap door" in the wall. But many defects could not be heard without playback, which is the subject of the *next* section.

Electrical recording permitted engineers to hear exactly what was coming through the microphone, but it seems to have been some time before the trap door system was abandoned and soundproof walls and/or double glazing substituted. The Western Electric system involved a loudspeaker with a direct radiating conical cone driven by a vibration mechanism exactly like the disc cutterhead. But this would not have been particularly loud, so there was no risk of a howl-round through the trap door. However, it was just sufficient for the engineer at the amplifier rack to hear the effects of altering the amplification, and many kinds of gross faults would be noticed when the trapdoor was shut. The fact that engineers could both monitor the recorded sound, and control what they were doing, was an enormous step forward.

Landline recordings from other locations permitted "infinite acoustic separation" between engineer and artists. In practice, there would be what we would now call a

"floor manager," although I do not know the contemporary term for the individual. This would be another engineer at the other end of the landline with the musicians, connected with the first engineer by telephone. The first engineer would guide the floor manager to position the artists and microphone by listening to his loudspeaker, and the floor manager would cue the first engineer to start and stop the recordings.

Photographs taken at the opening of the Abbey Road studios in 1931 show the trapdoor system was still in use, although the engineer would have needed lungs of iron to shout down a symphony orchestra at the other end of Studio 1! It seems to have been about 1932 or 1933 before double-glazed windows were introduced, together with "loudspeaker talkback." Now engineers heard what was coming from the studio with even greater clarity than the public. Although sound mixing definitely antedated this, especially in broadcasting, the new listening conditions enabled multi mike techniques to be used properly for commercial recordings for the first time. I could now write many thousands of words about the engineer's ability to "play God" and do the conductor's job; but if our objective copy is defined as carrying "the *intended* original sound," we shall not need to take this into account.

In the world of radio broadcasting, a different "culture" grew up. Broadcast monitor areas were conventionally like medium-sized sitting rooms. Broadcasting was considered an "immediate" medium. In Europe, anyway, broadcasts were considered better when they were "live," and this was certainly true until the mid-1950s. Thus the principal use of recording technology was so complete programmes could be repeated. It was therefore considered more important for a recording to be checked as it was being made, rather than being kept in a virgin state for mass production. In Europe three media were developed in the early 1930s to meet this need - magnetic recording on steel tape, nitrate discs, and Philips-Miller (mechanical, not photographic) film. With each of these media, engineers could, at the throw of a switch, compare the "line-in" signal with the "reproduced" signal. (Of course, this became practically universal with magnetic tape).

This meant the recording engineer usually had to be in a separate room from the engineer controlling the sound, because the monitoring was delayed by the recording and reproducing mechanisms. "Recording channels" comprising at least two machines and a linking console were set up in acoustically separate rooms. This also kept noises such as the hiss of a swarf pipe away. And, as the recording equipment became more intricate and standards rose, the art of sound recording split into two cultures - one to *control* the sound, and one to *record* it.

This split remained until magnetic tape recording equipment became reliable enough for the sound-balancing engineer to start the machines himself using push buttons. Then, in the 1960s, the cultures grew closer again. Tape recorders found their way back into the studio monitoring areas, tape editing could be done in close cooperation with the artists, and the "tape op" was often a trainee balance engineer.

13.9 The effects of playback

Although the engineer might play two rôles - a rôle in the recorded performance and a rôle in the technical quality - the former consideration was shared by the artist himself when it became possible to hear himself back. Nowadays, we are so used to hearing our own voices from recordings that it is difficult for us to recall how totally shocked most adults were when they heard themselves for the first time - whether they were speakers, singers, or instrumentalists. This capability played an important part in establishing what

we hear from old recordings nowadays. But, important though it may be, it is rather more difficult to ascertain *what* part it played! So in this section, I shall confine myself to a brief history of the process, and hopefully this will inspire someone else to write about the psychology.

Early cylinder phonographs were capable of playing their recordings almost as soon as they were completed, but there seems to have been little long term effect on the performances. The basic power-bandwidth product was very poor; so the artistic qualities of the recording would have been smothered by massive technical defects. Also the phonograph was (in those days) a new technology interesting in itself, rather than being thought of as a transparent carrier of a performance. Finally the limited playing time meant it was considered a toy by most serious artists, who therefore did not give cylinders the care and attention they bestowed on later media.

Early disc records, made on acid etched zinc, could be played back after ten or twenty minutes of development in an acid bath. Gaisberg recalled that artists would eagerly await the development process (Ref. 14); but it seems the acid baths did not work fast enough for more than one trial recording per session. Thus, it may be assumed that the playback process did not have an effect significantly more profound than cylinder playback.

When wax disc recording became the norm around 1901, the signal-to-noise ratio improved markedly. Almost immediately the running time increased as well, and VIP artists started to visit the studios regularly. Thus playback suddenly became more important; yet the soft wax would be damaged by one play with a conventional soundbox. The experts had to tell artists that if a "take" were played back for any reason, the performance would have to be done all over again. Trial recordings for establishing musical balance and the effects of different horns or diaphragms might be carried out, but the good take couldn't be heard until it had been processed and pressed at the factory. The factory might be hundreds of miles away, and the time for processing was normally a week or two. So it can be assumed that the artists had relatively little chance to refine their recorded performances by hearing themselves back. Indeed, anecdotal evidence suggests almost the opposite. Artists were so upset by the hostile conditions of a recording session, that the facility was principally used for demonstrating why they had to perform triple forte with tuba accompaniment whilst being manhandled.

Two kinds of wax were eventually developed for disc recording, a hard wax which was relatively rugged for immediate playback on a fairly normal gramophone, and a soft wax which had much less surface noise and was kept unplayed for processing. (Ref. 15). But with electrical recordings, meaningful playback at last became possible. The soft-wax and hard-wax versions could be recorded at the same time from the same microphone(s). Furthermore, Western Electric introduced a lightweight pickup especially for playing waxes (Ref. 16). It was reported that "A record may be played a number of times without great injury. At low frequencies there is little change and at the higher frequencies a loss of about 2TU" (i.e. two decibels) "per playing." But clearly this wasn't good enough for commercial recording, and it is quite certain that soft waxes were never played before they were processed. (See also Ref. 17). The decision was then swayed by considerations like the time wasted, the costs of *two* Western Electric systems (one was expensive enough), the cost of the waxes (thirty shillings each), and so on. Frankly, playback was a luxury; but for VIP artists at least, it was sometimes done. It seems principally to have been used in film studios; Reference 18 makes it clear that actors always crowded into a playback room to hear themselves, and modified their performances accordingly.

We saw in the previous section that instantaneous playback was an essential requirement for a broadcasting station, so from about 1932 onwards we may assume that playback to the artists *might* have been possible at any time. However, we can also see that it was very unlikely in certain situations - mastering on photographic film being the obvious case. Also, many broadcast recordings were made off transmission for a subsequent repeat; if the artist heard himself at all, it was too late to correct it anyway.

But I mention all this for the following reason. Let us suppose, for the sake of argument, that a VIP pianist heard a trial recording made with a Western Electric microphone. Would he have noticed that the notes around 2.9kHz were recorded 7dB louder than the others, and would he have compensated during the actual take? I stress this is a rhetorical question. There is no evidence this actually happened. In any case, the defects of the Western Electric moving-iron loudspeaker cone were greater than 7dB, and moving the microphone by six inches with respect to the piano would cause more variation than that as well. But here we have an important point of principle. Whenever playback of imperfect recordings *might* have taken place, today we *must* either make separate archive, objective, and service copies, or document exactly what we have done to the archive copy! It is the only way that subsequent generations can be given a chance to hear an artist the way *he* would have expected, as opposed to the *actual* sound arriving at the microphone.

The characteristics of contemporary monitoring equipment are not important in this context. Although a loudspeaker might have enormous colouration by present-day standards, it would colour the surface noise in such a way that listeners could instinctively reject its effects on the music. Certainly it seems people can "listen through" a loudspeaker, and this is more highly developed when the listeners are familiar with the system, and when it's a good system as well. Thus the judgements of professional engineers may dominate. In any case, the issue does not arise if there was a post-production process - a film doing through a dubbing theatre, or a multitrack tape undergoing "reduction" - because that is the opportunity to put matters right if they are wrong. It is also less likely when trial recordings were made and played back immediately with essentially flat responses. The writer's experience is that, when the artist showed any interest in technical matters at all, considerations concerning the balance arising from microphone positioning dominated. Finally, when the artist heard test pressings - particularly at home, where the peculiarities of his reproducing equipment would be familiar to him - the issue does not arise.

Although I was unable to find an example to back up my hypothetical question about the Western Electric microphone, there is ample subjective evidence that this was a real issue in areas where lower quality recording equipment predominated. It is particularly noticeable on discs recorded with a high-resistance cutterhead. In chapter 5 we saw this resulted in mid-frequencies being emphasised at the expense of the bass and treble. We can often establish the peak frequency very accurately and equalise it. But when we do, we often find that the resulting sound lacks "bite" or "sparkle." The engineers were playing test pressings on conventional gramophones whose quality was familiar from the better records of their competitors, and they relocated the musicians or rearranged the music to give an acceptable balance when played back under these conditions. You can see that potentially it is a very complex issue. It was these complexities which caused the most passionate debate at the British Library Sound Archive, and made me decide to write this manual to ventilate them. I have mentioned my solution (three copies); but it is the only way I can see of not distorting the evidence for future generations.

13.10 The Costs of recording, making copies, and playback

Clearly, these are further features we must take into account when we study the interaction between performers and engineers. But the trouble with this section is that money keeps changing its value, while different currencies are always changing their relationships. A few years ago I wrote an account of how the prices of British commercial sound records changed over the years (Ref. 19). I attempted to make meaningful comparisons using the official British "cost of living index", which (when it commenced in July 1914) represented the price of necessities to the average British "working man". By definition, this was someone paid weekly in the form of "wages". It automatically excluded people who received "salaries" (usually received quarterly in arrears), so their influence had no effect upon the "cost of living index". Yet for several decades, only salaried people could afford sound recordings.

I faced yet more financial problems in the original article, but here I shall not let them derail me - the effects upon this section are not very significant. So I am taking the liberty of dealing with costs from a purely British viewpoint, and because neither Britain nor America experienced the runaway inflation suffered in many other countries, I have simply decided to stick with my original method. I shall just express costs in "equivalent present-day pounds sterling" (actually for the end of the year 2000), and leave readers to translate those figures into their own currencies if they wish.

But I had better explain that British prices were quoted in "pounds shillings and pence" until 1971. (There were 12 pence to a shilling, and 20 shillings to a pound; and the symbol for pence was "d".). In 1971 Britain changed to decimal coinage with 100 "new pence" to the pound; but the pound stayed the same, and forms the basis for my comparisons.

The next dimension is that we should differentiate between the costs of the recording media (and the overheads of recording them), from the prices people paid out of their pockets for any mass produced end results. This was particularly important between the wars, when making satisfactory disc masters cost astronomical sums. This was only workable because mass production and better distribution made the end results affordable; but I shall study this dimension, because it radically affected what became recorded.

Two more dimensions are what the purchaser received for his money - first playing time, and second sound quality. For the former I have picked a "nominal playing time," about the average for records of that period. But I shan't say much about the quality (or power-bandwidth product), except to note when it altered sound recording history.

And there is a whole spectrum for the end results, from vocal sextets in Grand Opera to giveaway "samplers." Unless I say otherwise, I shall assume the "mainstream" issues in a popular series - in HMV language, the "plum label" range - rather than "red label and higher", or "magenta label and lower."

13.10.1 The dawn of sound recording

There is no doubt that news of Edison's invention stimulated heavy demand for music in the home, but this demand was fanned to blazing point by three unplanned circumstances. First, Edison's 1877 Tinfoil phonograph was not designed to allow

recordings to be changed. After the tinfoil had been unwrapped from the mandrel, it was practically impossible to put it back again on the machine which recorded it, let alone a different machine.

Bell and Tainter's "Graphophone" was the first machine to allow the medium to be changed, and Edison soon adopted the principle; yet I don't know any written history which even mentions this breakthrough! The next circumstance was that both the subsequent "Graphophones" and "Improved Phonographs" were targeted at businessmen for dictation purposes; the idea of using them for entertainment was resisted by nearly everyone in the trade. Finally, the various organisations marketing sound recording in North America and overseas got themselves into a horrible entanglement of permissions and licenses, which made it practically impossible to decide what was legal and what was not - even if anyone could have seen the whole picture, which nobody did.

After the first Edison wax cylinder phonographs began to fail as dictating machines, it was possible to buy one for a hundred and fifty dollars, which translates into £31. 5s. 0d. in British money of the time. When you take inflation into account, it makes no less than £2000 today. In his book "Edison Phonograph - the British Connection," Frank Andrews describes the early days of pre-recorded entertainment in Britain. It seems this did not start until 1893, and then only by accident. "The J. L. Young Manufacturing Company" was marketing Edison dictation phonographs imported from America. Although the Edison Bell Phonograph Corporation of London had purchased twenty Letters Patent on the use of phonographs and graphophones (to give them a legal monopoly in Britain), Young's phonographs bore a plate from the Edison works stating that their sale was unrestricted except for the state of New Jersey. With typical American xenophobia, the designer of the plate had forgotten the Edison Bell monopoly in Britain.

The Earl of Winchelsea was at Young's premises buying some blank cylinders for dictation purposes. The music hall singer Charles Coburn also happened to be trying a phonograph by singing into it, and the Earl overheard him and asked if the cylinder was for sale. Young sold it to him, and apparently Coburn carried on all day singing unaccompanied into the machine. Other customers were informed they were for sale, and the whole stock was cleared by 11 o'clock next morning. Unfortunately, history doesn't relate how much they fetched.

The first case where we know the price is also a case we cannot date - told by Percy Willis to Olgivie Mitchell (author of "Talking Machines", a book published in 1922). Willis admitted he had been breaking American law, since he was the first to smuggle a phonograph out of America to Ireland in about 1892, where he made 200 pounds in five days just by exhibiting it. He returned to the States for more machines, hid them under fruit in apple barrels, and recorded many cylinders which he sold at a pound apiece. Thus the cost was one pound for three minutes or so - about £64 when you take inflation into account.

Wax cylinder equipment was maintained for specialist purposes for another four decades, until electrical disc recording displaced it. Besides the "family records" made by phonograph owners, the format was widely used for collecting traditional music and dialect. It was powered by clockwork, so it could be used in the field without any power supply problems. A blank cylinder weighed just short of 100 grams. I do not know its original price, but it would have been about a shilling in 1906 (about three pounds today), and it could be shaved (erased) a dozen times or more - even on location. This made the medium ideal for traditional music and dialect until the 1940s. Allowing for inflation, and the cost of repairs and new cutting styli, call it six pounds for two minutes.

13.10.2 Mass produced cylinders

Between 1892 and 1902 the price of a one-off commercial cylinder had fallen as low as 1/6d - the best part of a fiver at today's prices, even though this was a "downmarket" make ("The Puck."). All these cylinders were one-offs, made by artists with enough stamina to work before several recording horns performing the same material over and over again - a circumstance which seriously restricted what we inherit today. In his book, Frank Andrews lists many other twists in the tale as other organisations tried to market cylinders. But Edison's two "tapered mandrel" patents in Britain expired in April 1902, while Edison's patent for moulding cylinders was disallowed; and after this, moulded cylinders could be marketed freely.

This launched the situation whereby a commercial sound recording tended to be higher quality than any other type of sound recording. This remained true throughout the eras of electrical recording, sound films, radio broadcasting, television broadcasting, and the compact disc. If the word "quality" is taken as meaning the artistic and craftsmanship aspects of sound recording *as well as* the technical ones, commercial sound recordings remained superior until the end of the twentieth century. It then became possible to purchase equipment with full power-bandwidth product, and add time, inspiration, and perspiration to create recordings which equalled the best.

Returning to the moulding process: this was comparatively inexpensive. To use a modern phrase, it was a "kitchen table" process; amortised over some thousands of cylinders, it can be ignored for our purposes. Edison and Columbia "two minute cylinders" never dropped below 1/6d until the "four-minute cylinder" took over, when the older ones became 1/-. Bargain two-minute cylinders eventually dropped as low as 8d. - equivalent to almost exactly two pounds today. Many cheap ones were "pirated" (to use twenty-first century vocabulary, although there was no copyright in sound recordings in Britain before 1912).

13.10.3 Coarsegroove disc mastering costs

The first disc records were mastered on acid etched zinc; but when the US Columbia and Victor companies formed their patent pool to allow discs to be mastered in wax, this considerably increased the costs of making original recordings. The slab of wax was about an inch thick (to prevent warpage and to give it enough strength to be transported to the factory for processing), and its diameter was greater than that of the manufactured record, so mothers and stampers could be separated without damage to the grooves. A wax for one ten-inch side weighed over 1.8kg. In other words, it comprised more than eighteen times as much wax as a blank cylinder, so it isn't surprising that master waxes were quoted as costing thirty shillings each in the mid-1930s (about £54 today), plus the cost of the machinery and operators to cut them, and of course the artists' fees!

Next came the "processing" - the galvanic work to grow masters, mothers and stampers. Here I do not know the actual costs (which were trade secrets). But in the mid-1960s when I was cutting microgroove on cellulose nitrate blanks, the first stamper cost about the same as the nitrate blank. So, ignoring the artists' fees, you might double that 1930s figure to £120 for three minutes - and this was before anyone could play the recording back!

In the circumstances, only professional performers who gave reliable performances would be normally be invited to make commercial records. This supported the idea of

different classes of artist, together with the idea of different colours of label for each of the different classes. When an artist's name sold the recording (rather than anything else), only one or two individuals with a talent for getting on with VIPs took on the job called "record producer" today. He and his artist would decide which version should be used *after* £120 had been expended making a "test pressing" of each take which might be publishable. And this is why *unpublished* takes seem only to survive for VIP performers.

But, before archivists fish in their pockets for wads of money, I must explain the test pressing system went further than that. It was quite normal for the *published* version to exist on quite literally dozens of so-called "test pressings", presumably for attracting new artists, or bribery purposes, or promotional reasons.

In the days before electronic amplification, the combination of acoustic recordings and devices like Pathé's mechanically magnified discs or The Auxetophone pneumatic amplifier (section 12.25), could make a utilitarian (if inflexible) public address system. Otherwise, only the wealthiest amateurs (facing a cost well into three figures by modern standards, without being upset by failures) could afford to make a sound recording until the 1930s.

13.10.4 Retail prices of coarsegroove pressings

Disc pressing was much more capital intensive than cylinder moulding. Apart from the presses themselves (which cylinders didn't need), there were ongoing costs for steam to heat the stampers, water to cool them, and hydraulic power for hundreds of tons of pressure. So for the first decade or more, disc records were even more expensive than one-off cylinders, let alone moulded cylinders.

Between 1902 and 1906, a mainstream seven-inch single-sided black label Gramophone disc retailed at 2/6d (playing time about two minutes), while the ten-inch single-sided equivalent (three minutes) cost five shillings (about £15.50 today).

In September 1912, "His Master's Voice" was actually the last label to market double-sided discs. They did so whilst ceasing to advertise in the Trade Press, which makes it difficult for me to establish how they were priced. (Clearly they were hoping not to alienate their loyal dealers, without at the same time getting egg on their faces). But it seems single-sided Gramophone Concert discs (then 3/6d) gave way to double-sided plum label B-prefix HMVs at the same price (equivalent to about £10.95 today). The table below shows what happened next.

The first line shows there was still room for price cuts, and was probably helped by the amortisation of the capital required to build the Hayes factory (construction of which began in 1907). After that, it is striking how consistent that right hand column is. There were comparatively small alterations in the untaxed costs of double-sided plum label discs, and the principal ones were independent of World Wars or anything Chancellors of the Exchequer could do.

I shall now deal with the "extremes", rather than "plum label" records. For many years, the most expensive "extreme" was considered to be Victor's 1908 recording of the sextet from *Lucia de Lammermoor*. Historian Roland Gelatt wrote that this was priced at seven dollars specifically for its ability to attract poorer classes of people to the store like a magnet. Its British equivalent cost fifteen shillings until it was deleted in 1946, equivalent to sixteen or seventeen pounds of today's money for four minutes of music. And it must also be said that here in Britain, this fifteen shilling issue was eclipsed by others by Tamagno and Melba.

HIS MASTER'S VOICE: B-series double-sided disc prices

Date (dd/mm/yy)	Retail price (shillings & pence, excl. tax)	Equivalent today
09/12	3/6d	c. £10.95
pre 06/17	2/6d	£4.00
03/18	3/0d	£3.98
09/18	3/6d	£4.47
??/19	4/0d	£4.64
12/23	3/0d	£3.99
mid.08/31	2/6d	£4.65
01/09/37	3/0d	£5.74
c.06/42	3/3d	£5.19
06/49	3/9d	£3.18

The cheapest records were retailed by Woolworth's. From 1932 to 1936 they were eight inches in diameter with the tradename "Eclipse", and from 1936 to 1939 they changed to nine inches with the tradename "Crown." At sixpence each, this is equivalent to about £1 today. This enormous range for retail prices underwrote both the prestige and the mass production of commercial disc recordings, in a way which never happened with cylinders.

At this point I must break off for two reasons. First, subsequent prices include Purchase Tax or Value Added Tax, and second, popular Long-Playing albums were now available. To show how these fared, I have averaged in the next table the Decca LK series and the HMV CLP series (which were always within a few pence of each other).

This time, it's apparent from both columns that, in "real terms", prices of records steadily fell. I leave it to readers to speculate why; but the result was that popular music grew to a billion dollar industry, with all the other genres easily outclassed.

It would be nice to bring this right up to date; but at this point Resale Price Maintenance was declared illegal. (British record companies had used it because royalties to music publishers were fixed at 6.25% of untaxed retail price by law. It would have been impossible to pay the correct royalties if retailers had *increased* their prices; but although they now gave a "recommended retail price", competition kept prices *down*, so music publishers got *more* than their fair share).

POP "SINGLE" and POPULAR 12" LP PRICES, including tax

Date (dd/mm/yy)	Single Price	Equivalent today	LP price	Equivalent today
06/49	4/8d	£4.03		
06/52	5/4½d	£4.40	35/0d	£27.55
08/53	5/0d	£3.72	32/4½d	£24.10
28/10/55	5/7d	£4.03	33/11½d	£24.39
01/59			35/10d	£23.83
08/65			32/2d	£17.10
04/67	7/5d	£3.75		
03/68			32/7d	£15.60
12/69	8/6d	£3.73		
12/70	9/6d	£3.69	39/11d	£15.50
03/72			£2.19	£14.68
12/72	50p	£3.06		
12/73	48p	£2.80		
12/74	55p	£2.60		
09/75			£2.62	£11.15
12/76	65p	£2.76		
12/80			£5.00	£10.64

13.10.5 One-off disc recording

In the mid-1930s the costs of wax mastering stimulated the development of "direct disc recording" on various formulations of blanks, the most familiar being "the acetate" (mostly cellulose nitrate). Every few months, the magazine *Wireless World* provided listings of all types of recording blanks. Double-sided ten-inch discs (six minutes) were around 2/6d each (about five pounds today), and twelve-inch (eight or nine minutes) were about 4/6d (about nine pounds). But unlike wax cylinders, they could not be shaved or "erased"; British sound recording enthusiasts simply had to record note perfect performances with rigorous cueing and timing, or pay the price penalty.

A side effect of this is that we often find discs with bits of the performance missing, and the exact opposite - sections of off-air recordings or "candid" eavesdroppings upon rehearsals, the like of which would never survive today. These prices had doubled by the end of the Second World War, becoming even more expensive than HMV "Red label" pressings. The discs were easily damaged as well. But the advent of microgroove fuelled a rebirth in the medium, since microgroove discs needed more careful handling anyway, and the blanks could now hold five times as much material. Linked to mastering on magnetic tape (with its editability), the strain on performers was largely eliminated. Unfortunately the prices of blanks continued to rise, passing those of manufactured LPs in about 1965; by 1972 the additional costs of the machinery, the cutters, and the engineers, meant cassette tape was better suited for one-off recording work.

13.10.6 Magnetic tape costs

For a number of reasons, disc was paramount in Britain for longer than anywhere else. When we consider open reel magnetic tape, the relationship between quality and quantity was unlike any other medium. If in 1960 you wanted to record a half-hour radio programme (for example), you might use a "professional" seven-inch reel of standard-play tape at 7.5 inches per second (19 cm/sec) "full track" at a cost of £2. 8s. 0d.; or you could use one of four tracks on a "domestic" seven-inch long-play tape at 3.75 inches per second costing exactly the same, but the reel would hold no less than twelve such programmes. Using these figures, one half-hour radio programme would cost the equivalent of from £20 to £1.67 today. Because the trades unions representing musicians and actors influenced what could be kept on broadcast tapes, broadcasting archives are now being forced to plunder such amateur tapes (often with detriment to the power-bandwidth product).

But by 1960 prices were already falling. The earliest definite price I've been able to find is £2. 14s. 0d. for a seven-inch long-play tape in July 1956, and it gradually became apparent that British-made tape was much more expensive than imported tape. Its cost fell about one-third between 1960 and 1970, during which time inflation rose almost exactly the same amount, after which British tapes were matching imported ones.

None of these figures include the capital cost of any tape recorders, let alone satisfactory microphones, studio equipment, and soundproofing. This capital element was easily the biggest component of mastering on tape for *commercial* sales. In the 1950s this element was at its height, and formed the principal reason why the British duopoly of EMI and Decca remained unbroken. But in the 1960s it became feasible for minor recording studios to stay afloat while making recordings of commercial quality (this element had happened a decade earlier in the USA), so the only extra things needed were mass production (pressing) and distribution! The former consideration was prominent in the 'sixties and 'seventies. Despite the commercial popular music industry being (arguably) at its height, it was very difficult to get time on a seven-inch press, so there are very few short runs of popular music pressings - only "acetates".

By 1970 the Philips cassette had been launched as well. A blank C60 (sixty minutes) cost just under a pound, while C90s were almost exactly £1. 5s. 0d, and C120s around £1. 13s. 0d. These were "ferric" cassettes, of course. Reels of tape did not attract purchase tax, because they were deemed "for professional use"; but when value added tax started in 1973, it went onto everything.

By 1980, new "non-ferric" formulations had doubled and trebled the prices, showing a demand for getting quarts of quality into pint pots. But ferric cassettes remained almost the same in pound notes, the only differences being the extra VAT. The secret of Philips' success was that Philips maintained strict licensing for the format, so cassettes would always be "downwards compatible" and playable on any machine. Cassettes didn't have different sizes, speeds, or track systems (continual difficulties with open reel tapes); they only had different magnetic particles.

"Cartridges" were an American invention. Cassettes proved popular in cars, since there was no pickup which could be shaken out of a groove and the tape was encased so it couldn't tangle. But American motorists favoured the cartridge because it could play continuously, without needing to be changed at awkward moments. Stereo and quadraphonic versions were offered; but they did not penetrate European markets very far, and the quadraphonic ones are almost unbelievably rare here.

13.10.7 Pre-recorded tapes and cassettes

By and large, pre-recorded tape media played second fiddle to analogue disc pressings; later, the same material might be purchased on three or four different formats. In Britain, EMI launched pre-recorded tapes in the autumn of 1954, and the reviewer in *The Gramophone* that November found it difficult to say anything praiseworthy except that you could tangle them round your leg and they would still play. They were 7.5ips mono half-track. The HMV HTC-series was the equivalent of a 12" plum label LP, retailing for £3. 13s. 6d. (more than double the LP). But this was a prelude to The Radio Show in August 1955, where EMI's "Stereosonic" Tapes were launched. This was the only stereo medium in Britain for the next three years.

Seven-inch spools of tape at 7.5 inches per second were featured. They consisted of one side of the equivalent LP copied from a magnetic master tape onto two "stacked" parallel tracks (so they couldn't exceed thirty minutes). A "plum label" selection cost £2. 4s. 0d (equivalent to £31.90 today). At this price reviewers insisted in documenting the playing time, which varied between eighteen-and-a-quarter minutes and twenty-two.

Stereophonic LPs arrived in mid-1958. Although there were attempts to charge more for stereo ones, most of the record industry put up with the inconvenience of "double inventory", and charged the same whether the LP was mono or stereo. By 1975 virtually all new microgroove releases were stereo, and virtually all users were equipped to play them (if not to *hear* them in stereo).

Open reel tapes always lacked a common standard which everyone could use for "downwards compatibility". But they had one supreme advantage which gave them a market niche - the complications of disc mastering and pressing were avoided. Therefore many organisations joined the recording industry to market pre-recorded tapes (and later cassettes) of unconventional subject matter. This would offer rich pickings for sound archives, were it not for their completely insignificant sales.

In 1971 the following "standards" were on offer:

2-track mono open-reel (3.75ips, 5-inch reel)
4-track mono open-reel (3.75ips, 5-inch and 7-inch reels)
4-track stereo open-reel (3.75ips or 7.5ips, 5-inch and 7-inch reels)
4-track stereo cartridge
8-track stereo cartridge
Philips cassettes (mono and stereo compatible)

In that year, Dolby's B-type noise reduction system made it possible to get gallons into pint pots. This was the fillip the Philips cassette needed. It became the first commercial rival to discs since the cylinder. Pre-recorded cassettes settled down at about 80% of LP prices, and this still continues today. One area - the talking book - has not succumbed to the blandishments of digital recording, and has even caused cassette duplicating plants to expand in this digital age. This was because the audiocassette had a unique feature - you could stop it, and start again from exactly where you stopped. Book publishers suddenly become cassette publishers; but even the original publishers have usually shortened the text, and cassettes are always dearer than the unshortened printed books. So I cannot compare like with like in this situation.

13.10.8 Popular music

After the 78rpm and 45rpm "popular single" formats I mentioned earlier, the industry spent much time flailing around trying other formats. Twelve-inch LP albums began to dominate, probably because of the sleeve artwork. The "seven-inch single" began dying about 1977, and was replaced by the "twelve-inch single". The price of crude oil (the basic material from which vinyl copolymer was made) dropped in real terms, so twelve-inch analogue single discs could be sold at much the same price as seven-inch, while the increased diameter meant noticeably cleaner quality (and louder music - the "louder is better" syndrome intruded yet again). Although this did not directly affect classical music, the same technology (and thinking) was used for a small number of twelve-inch singles of classical music (and promotional versions of classical music), until digital techniques came to the rescue.

The term "disc jockey" had first appeared in radio broadcasting. Disc jockeys had begun making regular appearances at dance halls in the early 1960s; but their principal reason for existence had always been to feel the mood of the audience, and select music to reinforce (or contrast) this mood. The twelve-inch single led directly to specific "disco" music. This was different from previous "popular songs"; it was danced to powerful loudspeakers which actually mesmerised the dancers, changing the course of popular music. It resulted in full time disc jockeys who worked in dance halls rather than radio stations, together with new artforms such as "scratching" and "rapping" (improvised vocals over a pre-recorded backing, using twelve-inch singles skilfully played one after the other without a break in the musical tempo. Compact discs always took second place in this environment, but manufacturers eventually got the prices low enough for CD singles to retail between £1.99 and £3.99.

13.10.9 Digital formats

We have now arrived at the period following Resale Price Maintenance. As I cannot have accurate figures, I will use my historical judgement. I shall simply state that digital compact discs (CDs) appeared in 1982.

As I remember they were about fifteen pounds each, about double the equivalent LPs, but offered three advantages which ensured success. First, they were much more difficult to damage; second, they had significantly greater power-bandwidth product; and thirdly they could have a rather longer playing time. The absolute maximum permitted by the "Red Book" specification is just short of 82 minutes, although the average is about seventy. This makes it practically impossible to compare like with like again, because LPs couldn't hold that much; and when performances were reissued, most record companies dutifully filled their CDs with extra tracks.

Today new releases from leading record companies still cost about fifteen pounds, but inflation has exactly doubled since 1982, so they are effectively half-price. And there are innumerable bargain issues at a fraction of this. Naxos CDs, for example, are offered at £4.99 (their cassettes are £4.49). There is a whole spectrum of other inexpensive material - many in double packs for the price of one.

But the costs of making CDs have fallen even more rapidly. Not only is it normal to give them away with magazines, but you can now buy a dedicated audio CD recorder for

only £250. If you're a computer buff, you can battle away and eventually do the same job at about half that price. And the blank discs are verging on fifty pence apiece.

13.10.10 Conclusion for pure sound recordings

This last thought shows that mechanical costs of making sound recordings are no longer an issue. Sound recordings are now at the point where costs aren't significant - straight marketing issues are the only considerations in distributing published sound recordings.

I'm now going to utter a sad thought, so readers who'd like a happy ending should read no further. The public can (and does) have access to great music, well performed, and recorded to standards which approach that of a healthy human ear. By and large they do not realise the technical, financial and artistic costs of getting to this point; so sound recording is no longer a "sexy" industry. It shows every sign of becoming like the water industry - you just turn a tap, and there it is. It is true films and video would be meaningless without it, but here too the resemblance to a water tap is becoming noticeable.

As I write this, music distribution via The Internet is beginning to conquer conventional retail sales. Quite frankly, successful lawsuits brought by the American commercial recording industry seem doomed to failure. The Internet is now developing alternatives to the "client server" model, so it will be impossible to say where any one piece of copyright sound recording is actually located. And the Internet community considers "free speech" to be a Human Right which has higher priority than rewarding multinational companies, so budding musicians put their works on the Internet on the basis "there is no such thing as bad publicity". Unless all the Berne Treaty countries unite quickly to reward creative people (by some process I cannot even imagine), the seventeenth-century metaphor of the artist in his garret (unable to bring his creations to the public) will once again mean the world in general will be deprived of the greatest art.

It'll take a major shortage to remind people about the cost of everything and the value of nothing.

13.11 The cinema and the performer

Before sound films, the only ways film actors could communicate meaning was by facial expression, gesture, obvious miming, and "intertitles" (the correct name for short pieces of film inserted to show what characters were saying). Personally, I would also add what is now called "body language"; and, judging by some of the ethereal writing of the "film theorist" community, obviously there's a Ph.D thesis here.

Academic readers of this manual will probably know about an extremely extensive literature on "film theory", and anyone not so aware might be recommended to start with Reference 20. However, nearly all sound engineers and directors working with pictures and sound have followed their practical common sense, rather than any academic theory. The literature is in fact extremely misleading for the purposes of understanding "the original intended sound", let alone its effects upon the performers (the subject of this chapter).

Silent film acting was of course supported by other arts, such as lighting, scenery, costumes, and editing (such as the grammar of "establishing shots" and "closeups"). During the filming, the director was able to speak to his actors and give running directions

throughout, and it was common practice to have a small instrumental ensemble to create a "mood" among the actors. The whole craft of silent film had developed into a very considerable art form with its own conventions, a classic case of performers adapting to a medium. This was almost totally overthrown when sound came along - for audiences as well as performers.

Prior to the "talkies," silent films might be accompanied by anything from a full symphony orchestra playing a closely written score, to a solo piano playing mood music to cues - according to legend, quite independently of the picture! Many cinemas often employed local workers with no special musical talent. When I was younger I was regaled with stories from a Roman Catholic uncle, whose choir was hired for suitable death scenes, and the choirboys improvised sound effects for the rest of the film. Picture palaces rarely had acoustics suitable for music, the projection apparatus was not always silenced, and I also understand it was quite normal for the audience to converse among themselves. Literate people would also read out the intertitles to those who couldn't read; and even *blind* people joined the audience, where it was both safe and sociable.

13.12 Film sound on disc

When sound films were first attempted with acoustic recording, it was almost impossible to get the artist where he could be filmed without also getting the recording horn into shot. The procedure was therefore to record the soundtrack first (usually at a commercial recording studio), process the disc, and then to "film to playback." Various arrangements of gears and belts might be needed to link the sprockets driving the film with the turntable; but essentially the subject matter could be no more flexible than acoustic recording techniques allowed - loud sounds running about four minutes. Often the same material would be pressed and sold as an ordinary commercial record, or the disc would subsequently become separated from the film. It isn't always apparent today when such discs were meant to accompany moving pictures. (Ref. 21). The next step was a backwards one for sound archivists. Film studios might shoot their artists "mute," and then pay them to sing or speak from the cinema orchestra pit, viewing the film in a mirror so they could direct the full power of their voices at the audience.

Early films would be in reels about one thousand feet long. The frame rate was not standardised until the advent of sound on film. Forgive me while I write about a side issue, but it underlines the message I wish to convey. When film speeds became standardised, cameramen lost a means of artistic expression. It is an oversimplification to assume that the films of the first quarter of this century always had comic actors charging around at express speeds. When we see the films correctly projected, we can also see that the cameraman might be varying the rate at which he turned the handle so as to underline the message of the film - quite subtly "undercranking" for scenes which needed excitement, or "overcranking" for emotional scenes. The speed might even vary subtly within a single shot. This is a tool which cameramen lost when sound began, and a means of expression was taken from them. So we must always remember that the relationship between a performance and the hardware is a symbiotic one.

So the introduction of sound caused a major revolution in cinema acoustics and cinema audiences, as well as the films themselves. Fortunately, Western Electric were at the forefront with their disc technology. I say "fortunately," because they alone had the know-how (however imperfect) to correct cinema acoustics. Although Western Electric's prices were very high, the associated contracting company Electrical Research Products

Inc. provided a great deal of acoustic advice and treatment. This usually comprised damping around and behind the screen to eliminate echo effects between the loudspeakers and the wall behind the stage, and sufficient extra damping in the auditorium to cut the reverberation time to be suitable for syllabic speech. It does not seem that modern methods of acoustic measurement were used; but I have found no written record that screen dialogue was too fast in a Western Electric cinema, so it seems the process was successful. At any rate, film makers never seem to have compromised the pace of their films for the cinema audience; and, by a process of "convergent evolution," this meant that *all* cinemas evolved to roughly similar standards.

The running time of a 1000-foot reel of silent film might be anywhere between fifteen and twenty minutes - much more than a 78rpm disc, anyway. Between 1926 and 1932 there were two different technologies for adding sound to pictures - disc sound and optical film sound. I shall start by considering disc sound.

Disc recordings switched to 33 1/3rpm to accommodate the extra duration. Electrical recording was essential because of the amplification problem. Because the technology introduced by Western Electric was more advanced, it played the leading role in the commercial cinema for a couple of years (roughly 1927-1929). It used the technology we described in section 6.15, with the addition of three-phase synchronous generator motor systems (called "selsyns") to link the camera(s) and the disc-cutting turntable(s). Unfortunately, it was impossible to edit discs with the ease of film, and film directors had to alter their techniques.

First came material which could be performed in one "take." In 1926 the vast majority of American film audiences were not being entertained in dedicated "picture palaces," but in multi-purpose entertainment halls in small towns. More than half the evening's entertainment preceded a silent feature film, usually including a small orchestra, singers, vaudeville artists, and film "shorts" with cartoons and newsreels. So it was natural for such live acts to be recorded with synchronous sound; and since both the subject matter and the place of performance already existed, they could be shot in one take. Meanwhile, there was enormous potential for reducing costs and improving artistic standards at the point of delivery. Also the process was much cheaper than feature film methods of production, and patents were less of a problem. The sound was recorded on disc in the usual way, while there might be one, two, or three cameras running from the same power supply. There would be a "master shot" (usually the widest angle camera) which would be continuous; then the film editor could substitute closeups or cutaways from the other camera(s) whilst keeping the overall length the same. The master disc would be processed and pressed in the usual way.

Vitaphone was the first successful film company to use synchronous sound filming. As Reference 22oints out, this process was lubricated because it was much cheaper to make a film of a vaudeville act and show it all round the country, than to pay the artists to tour the country! What is now accepted as being the first "talkie" ("The Jazz Singer", starring Al Jolson) was initially shot using this system, with the specific idea that if it failed as a "talkie", it would still be possible to use the material in musical "shorts" (Ref. 23). And it was not a *complete* talkie; the film included several reels with "silent film music" on disc. Its fame is probably due to an accident - a piece of ad libbed speech by Al Jolson - which generated the very idea of "talkies", and happened to be followed almost immediately by a disc of droning silent film music. Vitaphone had also developed their technology specifically to save costs of hiring musicians to accompany silent films; the first full length movie with a continuous music soundtrack was recorded this way and

premiered in 1926 - "Don Juan." The full story of how thirty years of film art was junked, while sound was economically inevitable, is given in Ref. 24.

There now followed a period of considerable confusion, when established Hollywood stars were junked because they had "unsuitable voices". For perhaps the first and only time in the whole history of sound recording, a sound engineer had a right of veto over the performance. To some extent this is understandable. The craft suddenly lurched forwards, to a stage analogous to live television a quarter of a century later. Everything happened in realtime - the dialogue, the singing, the action, the cutaways, the camera moves and the lighting changes, all completed with a live orchestra off the set. It is not difficult to tell when this happened, because there will be long sequences shot "television fashion" with several cameras operating at once. The performers will likewise be playing "safe", and projecting their voices because neither the camera nor the microphone could get in for tight closeups. The microphone had a 2.9kHz peak, as we saw in section 6.5.

The only way of editing sound while it was on disc was to go through another generation, using radio broadcasting techniques to cue the discs. The picture would then be "fine-cut" to match the discs. Reference 25 says the first film to be put through this process ("Old San Francisco") had audibly degraded sound, as a result of efforts to add earthquake sound effects (presumably to a musical score). The Marx Brothers polished their performances on the legitimate stage before they went to the film studio; this ensured they could get through many minutes of slapstick action without mistakes, while they could time their act so the cinema audience could laugh without drowning dialogue. Until modern technology came to its rescue, the second Marx Brothers film "Animal Crackers" (1930) - recorded on disc - suffered a major picture editing problem when Captain Spalding appeared to get out of his palanquin *twice*. This is a classic case where there are two answers to the "subjectivism" problem I mentioned at the beginning of this chapter. What would the original engineers have defined as "the original *intended* sound"? Would they have preferred us to use the modern tidied-up version, in which a couple of bars of music were cut and spliced while maintaining the rhythm; or would it be better to keep the faulty picture edit to underline the deficiencies of the disc system, while retaining the music as the composer wrote it?

Because early sound-on-disc films gave cinema projectionists difficulties (it might take some seconds for the lumbering mechanism to get up to speed), it became the practice for the end of each reel to finish with a shot of an artist staring fixedly and silently at the camera for ten seconds or so, while the disc comprised unmodulated grooves. This gave the projectionists time to switch over as and when the next reel was running without a visible gap. This gives difficulties today. Ideally, the "archive copy" should have the zombie-shot at full length, while the "service copy" should have the action tightened up to minimise the hiatus.

Vaudeville halls continued in action. They later filled a long felt want during the Depression, when a sound film programme could provide a couple of hours of escapism from the troubles.

13.13 The art of film sound

The vital years of development for the new art form were 1929 to 1939. Sound-on-Film offered better prospects to the feature film industry, and RCA Photophone did the basic research and development as fast as they could. Unfortunately, they were about a year

behind Western Electric, and many cinemas were equipped with the disc system. However, the ability to cut and splice the sound as well as the pictures eventually tipped the balance so far as *making* the films was concerned, and disc mastering virtually ceased after 1930. Before that, it is my belief that the major achievement was to get anything audible at all, rather than to impose any "art" upon it. I choose the year 1929 as being the turning point, because several things happened that year.

Firstly, microphone "booms" and "fishpoles" were adopted, allowing film actors to move freely around the set (instead of being fixed to the spot to address a microphone hidden in a flower vase). Secondly it became possible to cut and splice optical soundtracks without a loud "thump", thanks to the invention of the "blooping" technique, where an opaque triangular patch over the splice resulted in a gradual fade-out and fade-up. Thirdly, the first true use of "audio art" occurred, when Hitchcock (in his film "Blackmail") used deliberate electronic emphasis of the word "knife" in some incidental out of shot dialogue, to underline the guilt of the heroine who had just killed a man with a knife.

Because silent films had nearly always been screened with live music (even if the tradition of an upright piano playing "Hearts and Flowers" is a gross caricature), it was perfectly understandable for early sound films to concentrate on music at the expense of speech; and when synchronous speech became feasible, there was a learning curve while stage techniques were tried and abandoned. By 1929, actors had settled down to what is now the convention - a delivery reminiscent of normal speech, but actually projected more forcefully (somewhat analogous to vocalists' "crooning" technique).

A few early movies slavishly kept the two components, sound and pictures, strictly together; we can see the effects in the first reel of "42nd Street" (1932), where a mechanical thump, a change in sound, and a change in picture happens every time a new actor says a line. But this is also the earliest film I have seen with a better soundtrack from a frequency response point of view; presumably it used one of RCA's early ribbon microphones. Soon, working with sound and pictures independently became the norm, and devices such as the sound "flashback" and "thinks sequence" became part of the film editor's vocabulary.

More important, it was possible to "dub" the film. In films, this word has two meanings. First, it means adding music and effects, and controlling the synchronous sound, to make an integrated artistic whole. Second, it can mean replacing the synchronous sound with words from another actor, or in another language. To achieve both these ends, the film industry was first to use "multitrack recording." All these facilities excite controversy when they go wrong; but the fact that thousands of millions of viewers have accepted them with a willing suspension of disbelief shows they have an enormous artistic validity. By 1933, the film "King Kong" had pictures and sound which were nothing remotely like reality, but which convinced the cinemagoers. It could never have been possible with disc recording.

Despite the advantage of sound-on-film, it took longer to change the cinemas. Fortunately, RCA Photophone was able to invent ways round Western Electric's patents, and they offered their equipment to cinemas at a fraction of Western Electric's prices. But disc equipped cinemas were still operating in Britain as late as 1935, the records being transfers from the optical film equivalents, so the boot was on the other foot. If we want to see a film today whose sound was mastered on disc, we usually have to run a magnetic or optical film copy, because synchronous disc replay equipment no longer exists. Yet we are often fighting for every scrap of power-bandwidth product. Today's sound restoration

operator must view the film, and decide whether it was mastered on disc; then select the appropriate format if he has any choice in the matter.

Meanwhile, a very confusing situation arose about sound film patents, which I won't detail here (it involved three American conglomerates and one German one). Although cross-licensing kept the film cameras running, the whole situation was not sorted out in law until the outbreak of the second World War, by which time the various film soundtrack types I mentioned in chapter 7 had all been developed. The standard position for optical soundtracks on 35mm "release prints" was also finalised by 1929.

The problem of foreign languages then arose. At first many films were simply shot in two or three languages by different actors on the same set. Reference 26 points out a different scenario, in which the comic duo Laurel and Hardy brought foreign elocutionists onto the set to teach them how to pronounce foreign languages "parrot fashion". The resulting distortions only added to audiences' hilarity! Meantime, the high cost of installing sound projection equipment meant most countries were a couple of years behind the United States, and the system of intertitles continued for some time.

13.14 Film sound editing and dubbing

I write this next paragraph for any naïeve readers who may "believe what they hear." As soon as it became possible to move away from disc to film sound, it also became possible to cut and splice the pictures independently of sound - and, indeed, make the "whole greater than the sum of the parts" by "laying" different sound, changing the timing of splices, and modifying sound in processes exactly analogous to various lab techniques for the pictures. Nearly every film or video viewer today is unaware of the hard work that has gone on behind the scenes. The man-days needed to create a convincing soundtrack often outnumber those for the picture; and I am now confining my remarks to just one language.

Since the normal film-sound conventions already allowed the use of two tracks ("live sound" from the set, and music), it was natural for the next split to be between speech and sound effects. Precisely how this became the norm does not seem to have been related anywhere (another Ph.D thesis, anyone?), but it allowed Paramount to set up a studio in Joinville (near Paris) specifically for adding foreign dialogue to American films. This began operations in the spring of 1930. The studio was made available to other Hollywood film companies, and although it closed a few years later, Paris remained a centre for the "foreign dubbing" industry.

Here I use the word "dub" to mean re-voicing an actor, using syllables matching the original lip movements in the picture. As this craft developed, original "stars" would sometimes re-voice their parts back in Hollywood, particularly on location shots where low background noise could not have been assured. (The technology came too late to save the careers of silent-screen lovers with squeaky voices, or noble-looking actors with uneducated accents). My next anecdote keeps resurfacing, so I will relate it now. The Italian film industry, knowing all its films would have to be dubbed, apparently did not ask its actors to learn lines at all. Instead, they apparently counted aloud, and the dubbing actors would interpolate some reasonable dialogue afterwards!

Thus, from 1930 onwards it was normal to have three soundtracks (even for later television films), "music", "effects", and "dialogue". These would often be on the same strip of celluloid, whether optically printed or magnetically recorded.

In section 13.4 above I considered the subject of "sound perspective." Film aficionados went through a massive debate on this subject in the 1930s. Here the problem was one of "naturalness" versus "intelligibility". Basically, the optical film soundtrack (even *with* "noise reduction", did not have enough dynamic range to allow speakers a hundred yards from the camera to *sound* a hundred yards from the camera. The problem was exacerbated by solid walls being made out of cardboard, inappropriate studio sets giving the wrong kind of footsteps, and other acoustic phenomena. In practice, sound recordists did the only sensible thing. They placed the microphone as reasonably close to the actors as they could (usually about a metre above their heads, but facing their faces, so fricatives and sibilants would be caught). They worked on the assumption that if a sound was meant to sound distant, this could then by simulated by volume controlling, filtering, and/or artificial reverberation in the final-mix dub. Generally this remains true to this day, even in television. Today, philosophers of film technique justify the general consistency of speech quality by saying there are always at least *two* people involved, the speaker and the spectator! (Ref. 27).

Another difficulty is "silence." The effects track should normally carry a continuous element, which glues the entire scene together. Scene-changes (and time-changes) can be signalled subliminally to the viewer by changing this continuous element. In my work as a film dubbing mixer, this continuous element was the hardest part of constructing a convincing soundtrack, because everything went wrong if it disappeared. (I suspect a great deal of film music might have been composed to overcome this difficulty)! Even "natural" sound picked up on the set includes subliminal components like camera noise and air conditioning, which stick out like a sore thumb when they are absent. Therefore, film performances may require additional background noise, which has to be moved or synthesised to cover mute reaction shots and the like. A professional sound recordist will always take a "buzz track" for every location, including at least a foot or two of film running through the camera. Generating "silence" which does not resemble a failure of the sound system is usually possible; but of course, inherently it adds to the sounds made by the performers.

On the other hand, it is my duty to record that several film directors have avoided mixing and dubbing film sound on ideological grounds, while most semi-professional and amateur video systems simply eliminate all chance to work creatively with sound at all.

In the next section I shall explain how the limiter was introduced to improve the clarity of speech (by making all syllables the same amplitude on the optical soundtrack). The "louder is always better" syndrome kept creeping in, both in the cinema and on television; and when Dolby noise reduction became available for cinemas, most simply raised the volume of the soundtrack to make it louder while not increasing the background noise. Therefore it became possible to plan a sound mix to be deafening. In "Star Wars" (1977) this fundamentally affected the dialogue, since the music and effects could be very exciting, and the dialogue was specifically written and performed so redundant lines might be "drowned." When this film moved to television (before Dolby sound was available to the home viewer), it was necessary to re-mix the three components to prevent viewers complaining they could not hear the dialogue. Comparing the cinema version with the domestic version reveals the deliberate redundancies in the script, which stick out like a sore thumb when made audible!

Meanwhile, when commercial television began in Britain in 1956, viewers began complaining about the loudness of the advertisements. These complaints continue to this day; but if you *measure* the sound (either with a peak-reading meter or an RMS meter), advertisements always read *lower* than the surrounding programmes. This demonstrates

that subject matter always contributes a significant psychological component to "loudness." The problem got worse as audio got better, mainly because the "trailers" advertising a forthcoming feature film always utilised the most exciting (and loudest) bits (and the same happened with television trailers). Messrs. Dolby Laboratories were forced to introduce a "movie annoyance meter" in order to protect their name; but Jim Slater of the British Kinematograph Sound and Television Society is quoted as saying: "If cinemas no longer turn the sound down to suit the trailers, they will play features louder. Not everyone will like that."

13.15 The automatic volume limiter

In the meantime, the film industry led the world in another technology - volume compression. We have dealt with the nuts and bolts of this in chapter 10, so now we will consider the effect upon performances.

Natural speech consists of syllables which vary in peak strength over a range of twelve decibels or more. When dialogue was reproduced in a cinema, it had to be *clear*. Clarity was more important than fidelity, and with the restricted dynamic range of the first optical soundtracks before "noise reduction", the "variable density" soundtrack had an advantage. It gave "soft clipping" to the vowel-sounds of words, allowing an unexpectedly loud syllable to be "compressed" as it was being recorded. Although considerable harmonic distortion was added, this was reduced by the "constant amplitude" characteristic employed by optical film sound (section 8.7), and the result was preferable to the unexpectedly loud syllable.

Anecdotal evidence explains how the loud sounds of gunshots were aided by driving the soundtrack further, into "peak clipping". I have no actual experience this was the case; but traditional movie sound for a revolver comprises a comparatively long burst of "white noise", the spectrum of which changes little with peak clipping. Anyone who has tried to record a real revolver with a real microphone onto a well engineered digital medium will know that the pulse of the shock wave dominates. Thus, the traditional revolver noise in film Westerns is very much a child of the medium upon which it was being recorded.

Both variable area and variable density soundtracks were made by light being fed between low mass aluminium ribbons in a "light valve". These ribbons were easily damaged by an overload, and I have no doubt that early amplifiers driving these ribbons were deliberately restricted in power to reduce the damage. Another anecdote tells how Walt Disney, who did the first voices for "Mickey Mouse", blew up a light valve when he coughed between two takes. Clearly there would have to be a better way.

Western Electric introduced the first "feed back limiter" (sections 11.6 and 11.7) in 1932. This could be used to even out syllables of speech and protect the light valves at the same time. From that point on, all optical sound recording machines seem to have had a limiter as an overload protector, and in my personal opinion this tool became overused. Not only did the three components (dialogue, music and effects) each have a limiter, but so did the final mix as heard in the cinema. In chapter 10 I explained why it may often be impossible to reverse the effects of a limiter; but at least we have the ethical advantage that the film company obviously intended the sound to be heard after two lots of limiting.

As I said, optical sound is recorded with constant amplitude characteristics, meaning that high frequencies are more liable to overload. When an actor had defective teeth (I understand teeth were not as healthy in those days!), the resulting whistles could

cause considerable overloads. Within a couple of years, the "de-essing" circuit was incorporated into limiters. The de-esser greatly reduced the overloads, and also improved the balance between consonants and vowels. Therefore virtually all optical film dialogue since about 1935 has been "distorted" in this manner.

Yet this is not the end of the difficulties. When an old film is "restored", the audio usually goes through yet *another* limiter! In this case, my view is that here the techniques of Chapter 10 *must* be invoked to "restore the original intended sound." Fortunately, I am not alone in this; at least some film enthusiasts support the idea of optical printing the sound, when another stage of limiting will not happen. Either way, the disadvantage is that background noise may double; only some of the techniques of Chapter 3 will be any help.

Because of the effects of the limiters (which compress the dynamic range at a syllabic rate), and because the sound has to be enormously amplified to fill a cinema with 2000 seats, various psychoacoustic distortions occur. These were compensated empirically until 1939, when all the psychoacoustic and physical phenomena were brought together in a seminal paper (Ref. 28). The result was that speech tracks were mixed using a standard frequency curve called "The Academy Curve" or "Dialog Equalization". This attenuated the bass and added a certain amount of extra energy in the 1kHz to 4kHz range. If you take a speech track from a film intended for screening in a cinema (this does not generally apply to television film soundtracks, although some made in established film studios may also have it), recovering the original sound may require the Academy Curve to be reversed.

An automatic volume limiter has also become a tool to identify "commentary" or "narration." In this situation, a voice describes something which is not apparent in pictures, while the viewer must not mistake it for something happening on-screen. This is achieved by a combination of several sound engineering tricks. First, the speech is delivered much closer to the microphone than synchronous speech, so it resembles someone sitting immediately next to the listener; second, it is delivered in a "dead" acoustic, so it gains an impersonal quality; thirdly, the limiter tends to be pressed beyond the 4dB to 8dB limit given in Chapter 10 as being the point where listeners unfamiliar with the original don't notice anything; and finally noise gating (the opposite of compression) may be applied to eliminate intakes of breath. All these modifications of "the original sound" tend to robotise the voice, helping it to be distinguished from something "in shot"; and I once created much puzzlement for a radio drama producer when I imported this technology for the narrator of a radio play.

13.16 Films, video, and the acoustic environment

You would have expected the film industry to have followed in the footsteps of the audio industry (section 13.4), but no. As post-production methods developed, the techniques of augmenting or replacing dialogue and sound effects grew at an incredible rate, but not the use of acoustics. This was probably because films were meant to be seen in relatively large auditoria. Although cinemas were "deader" than the average speech meeting hall, let alone a concert hall, their acoustics were always considerably "liver" than the relatively subtle differences between (say) a kitchen and a sitting room. So these differences were not emulated on the film soundtrack.

By 1933 it was possible to film dialogue in the teeth of a howling gale (e.g. a studio wind machine). This did not need sophisticated windshields or filters. At first,

directors attempted to save money by shooting without sound and getting actors to voice the sequences later; but it was soon found better to use a distorted recording to guide the actor. So it suddenly became possible to film action on location, even before directional microphones were invented. But some studios also offered the facility to their top "stars" so they might have several tries at their lines without the need for picture retakes. This explains why, when replayed today in room with a short reverberation time (or seen on TV), "star" actors sometimes sound if they are in a different place from the rest of the cast.

Another difficulty was to get the actors to pitch their voices correctly in a quiet studio. Indeed, Reference 29 shows that, in the 1930s, different film studios had different practices in this matter. But when lines are being "redubbed", for example when simulating delivery in the teeth of a gale, the engineers quickly learnt the advantages of driving the volume of the background sounds to the actor's headphones, so the voice would be projected by the right amount for the scene being depicted. But, equally, this writer gets annoyed at hearing V.I.P. actors who obviously aren't aware of what is supposed to be going on around them.

13.17 Making continuous performances

Continuous performances of long films pre-dated sound. Despite being made on 1000-foot reels, all commercial films were made on the assumption that they would be shown in cinemas with two projectors and staff to ensure continuous pictures. When semi-professional formats (like 16mm film) came on the scene, this rule still applied. Amateur cine enthusiasts dreamt of emulating a commercial cinema in their living rooms, and some actually achieved it. However, 16mm film could be in 2000-foot reels, and as it ran at 40% of the speed, running times could be as much as fifty minutes. Commercial feature films were transferred to narrow-gauge formats for wealthy amateurs, and were also shown in "third world" locations, aircraft, church halls, and the like. They were usually on two reels, with no alteration to the way the film was presented. Thus it is conventional for us to simply join the reels up when we transfer them to new media today. The format was also used for instructional and documentary films, but hardly any of them were as long as this.

13.18 Audio monitoring for visual media

In early film studios, the sound engineer on a "sound stage" worked at a small wheeled console on the studio floor. This helped communications with lighting staff, microphone boom operator, etc.; but the basic sound monitoring was on headphones. The signal was sent by cable to a fixed sound camera room, where another engineer looked after a film recording machine. It seems it was this second engineer who was responsible for the limiter which protected the light valve. Presumably he could communicate with the engineer on the floor if anything sounded wrong, for instance if he heard infrasonic wind noise; but I have no evidence this actually happened.

From the earliest days, the second engineer could also monitor the performance of the limiter and light valve by a photo-electric cell taking some of the light. The light might be diverted by a partially-silvered mirror before it fell on the film, or it might be located *behind* the film so the 4% of light not absorbed in the emulsion would be used. The

switching was arranged so the loudspeaker was on "line in" before a take, and "PEC Monitor" when the sound camera was started. This monitoring was provided as an equipment check rather than an artistic check. This monitoring point is also known to have been used as a suitable take-off point for other purposes. For example, in the late 1940s and early 1950s, quarter-inch magnetic tapes were taken for commercial records of film soundtrack music. These would therefore have been through the limiter and light valve, but not the optical film itself. (And a very misleading convention developed, whereby an album called "the Original Motion Picture Soundtrack" actually means "the Original Motion Picture *Music* - after rejecting unused takes").

We also know that orchestral music for films was monitored in the conventional manner for sound only media, with the first engineer working with loudspeakers in a soundproof listening area. This would be slightly "deader" than the "theatre optimum", so the effects of natural reverberation from the scoring stage could be assessed. But in all other film monitoring processes (including dialogue re-recording, dubbing, and reviewing), the normal practice was to monitor in a "theatre optimum" environment, using a listening area designed to emulate a cinema-sized auditorium.

The American Standards Authority established national standards for indoor theatres. It is interesting that many domestic "Dolby Stereo" decoders can supply artificial reverberation to simulate a cinema auditorium in the home. Whether this is a chicken or egg situation is difficult to say!

We must also remember that engineers often made recordings specifically for "internal" purposes, i.e. not for the public. It was normal practice to have an analogue audio tape recorder continually running in a television studio, for several reasons. For example, it might provide instantaneous checks on the lines of "Was the soloist flat in Verse 2?" It might provide the elusive "background atmosphere" so vital in post-production, and which was quite inaudible next to a whirring video machine. It could supply alternative sounds which could be fitted after video editing had taken place, or separate audience responses for various doubtful motives, and to provide a tape with better power-bandwidth product than the video recorder for commercial LP releases. These tapes are known in Britain as "snoop tapes." They were "twin-track" from a comparatively early date (the mid-1960s), and often have SMPTE/EBU timecode in the guard band. They are neither "mono" nor "stereo." They provide a fertile hunting ground for students of television history.

13.19 Listening conditions and the target audience and its equipment

This brings us to another "cultural" issue - assumptions made by engineers about the way sound was *meant* to be reproduced.

Until the mid-1920s it can be presumed that, when they thought about the matter at all, engineers would assume their recordings would be played on the top of the range machine made by their employer in a domestic listening environment. It would also be assumed that the playback would not last much more than four minutes without a break, and this break would be sufficiently long to allow needles to be changed, clockwork motors to be wound, etc.

Broadcasting was the medium which pioneered "freedom from interruption." From day one, British broadcast engineers were taught that "the show must go on," and anything which interrupted a continuous transmission was a disciplinary offence. This signalled the start of background or "wallpaper" music in the home. It had been normal

in restaurants and the like for many decades, but now there was a demand for it in the home. Suddenly, there were two audiences for broadcasts (and almost immediately records). First there was the "serious listener," who could be assumed to be paying attention and listening in reasonable conditions. Secondly there was the "background listener," who might be doing several other things at once, and in a noisy situation. Frankly, engineers always worked on the first of these two assumptions until at least the 1950s. But, reading between the lines, it is also clear that more than half the population would be in the latter group. Printed evidence lists extensive complaints about the relative volumes of music and announcements, the unintelligibility of speech, the disturbing nature of applause or laughter, and the ubiquity of radio reproduction in public places. In Britain we have many recordings of radio programmes dating from the same times as such complaints, and (judging by our reaction today, admittedly) it seems the serious listener would not have been concerned. But we must also remember that until 1942 British broadcasting had no standard metering system, limiters, or network continuity suites; so judging one single radio programme on its own isn't a fair test. As we usually have only isolated radio programmes from this era, we can assume that anyone who wants to hear it will be paying full attention.

At this time, there was only A.M. (amplitude modulated) radio. Because it had to serve both "serious" and "background" listeners, there was a tendency (in Europe, anyway) for A.M radio to be broadcast with the minimum possible dynamic correction and the maximum frequency response. In 1960 this writer visited the BBC Home Service Continuity Suite in Broadcasting House where there was a Quad A.M Check Receiver for monitoring London's A.M transmitter at Brookman's Park. It was quite impossible to hear any difference between the studio sound and the transmitted sound, even on A.M radio. Surviving nitrate discs made off-air for broadcasting artists from the mid-1930s onwards show extraordinary fidelity for the time. But when the American Edwin Armstrong invented the F.M. (frequency modulation) system in the mid-1930s, it was suddenly possible for the two uses of radio to be split. Economic conditions made this slower in Europe, but by 1958 F.M. Radio had come to most parts of Britain, duplicating what was going out on A.M. The Copenhagen Plan for A.M. Broadcasting restricted the frequency range of A.M. Radio to 4.5kHz to prevent mutual interference. Thus serious listeners switched to F.M. (which also had lower background noise), and car radios tended to be A.M. because it was less susceptible to "fading" when driving in valleys or between high buildings. In consequence of this, broadcasters also tended to compress the dynamic range of A.M. so it would be audible above the car engine noise, and speech processing kept up the intelligibility in the absence of high frequencies. Before long radio broadcasters were targeting their transmissions accordingly, and we must remember the effects upon off-air recordings surviving today.

The supposed target audience frequently affected the "original intended sound." In chapter 10 we saw how the "Optimod" unit and similar devices reduce the dynamic range for car radios. The use of some form of compressor was virtually essential for most cinema and television films, because the sound mixer had to wedge three layers of sound (dialogue, music, and effects) into a dynamic range of only about 20dB. And we have also seen how excessive vertical modulation could restrict what went onto a stereo LP. But we now have a few compact disc reissues, where the stereo image is suddenly wider for low-pitched instruments of the standard orchestral layout. I am sure readers of this manual will be able to think of examples like that when they read almost any page. The basic fact is that the "medium" and the "message" are usually inextricably linked.

13.20 The influence of naturalism

Roland Gelatt has pointed out that the idea of a sound recording being a faithful representation of the original is a chimera running through a hundred years of history. It is true that many recording engineers (including those of Edison) conducted tests seeking to prove this point to sceptical consumers, and that they were often successful. It is also true that no-one with the ability to monitor sound in the manner of section 13.5 above has been able to resist running to the room with the performers and seeing what it sounds like there. But it must be confessed that the results of these experiments haven't been very significant.

The fact is that sound recording is an art form, not a craft. Its significance as a faithful preserver of sound is usually no greater than that of the film "King Kong" documenting the architecture of New York. Whilst sound archivists clearly have a responsibility to "preserve the original sound" where it exists, their responsibilities to "preserve the original *intended* sound" are much greater. Let us close this chapter, and this manual, with our thanks and respects to the engineers who elevated sound recording from a craft to an art.

REFERENCES

1: Peter Adamson, "Long-playing 78s" (*article*). The Historic Record, No. 17 (October 1990), pp. 6 - 9.

2: The Gramophone Company (in Germany), matrixes 1249s to 1256s. Reissued on CD in 1991 by Symposium Records (catalogue number 1087, the first of a double-CD set).

3: Percy Scholes, "The First Book of the Gramophone Record" (*book*), second edition (1927) page 87, London: Oxford University Press.

4: "Impact of the recording industry on Hindustani classical music in the last hundred years" (*paper*), by Suman Ghosh. Presented at the IASA Conference, September 1999, Vienna, and reprinted in the IASA Journal No. 15, June 2000, pages 12 to 16.

5: Peter Copeland, "The First Electrical Recording" (*article*). The Historic Record, No. 14 (January 1990), page 26.

6: In America, issued on Columbia 50013D, 50016D, and 348D. In Britain, only four titles were issued, on Columbia 9048 and 9063.

7: Gramophone 1462 or His Master's Voice E.333.

8: Sir Compton Mackenzie, "December Records and a few Remarks," The Gramophone, Vol. 3 No. 8 (January 1926), p. 349. Quoted in Gelatt: "The Fabulous Phonograph," (2nd edition), London, Cassell & Co., p. 232.

9: Val Gielgud and Holt Marvell, "Death at Broadcasting House" (*book*). London: Rich & Cowan Ltd., 1934.

10: Alec Nisbett: "The Technique of the Sound Studio" (*book*). London: The Focal Press (four editions)

11: (reverb plate)

12: R. F. Wilmut, "Kindly Leave The Stage!" (*book*), London: Methuen, pp. 68-9.

13: Victor 35924 or His Master's Voice C.1564 (*78rpm disc*), recorded Liederkranz Hall New York, 21st May 1928.

14: F. W. Gaisberg, "Music On Record" (*book*), London, 1947, pages 33 and 44.

15: P. G. A. H. Voigt, "Letter To The Editor," *Wireless World* Vol. 63 No. 8 (August 1957), pp. 371-2.
16: Halsey A. Frederick, "Recent Advances in Wax Recording." One of a series of papers first presented to the Society of Motion Picture Engineers in September 1928. The information was then printed as one of a series of articles in Bell Laboratories Record for November 1928.
17: Jerrold Northrop Moore, "Elgar On Record" (*book*), London, EMI Records Ltd., 1974. Pages 70-72 tell the story of an attempt to record Elgar's voice during an orchestral rehearsal in 1927. The wax was played back twice on the day, and then Elgar asked for it to be processed. The resulting pressings have given difficulties when played today; many of Elgar's words have been lost.
18: Scott Eyman, "The Speed of Sound - Hollywood and the Talking Revolution" (*book*), New York: Simon & Schuster 1997, page 184.
19: Peter Copeland, "On Recording the Six Senses" (*article*), Sheffield: The Historic Record, No. 11 (March 1989), pp. 34-35
20: (ed. Elisabeth Weis and John Belton): Film Sound Theory and Practice (*book*), Columbia University Press (1985).
21: Baynham Honri, "Cavalcade of the Sound Film," based on a lecture given to the British Sound Recording Association on 20th November 1953 and published in the BSRA Journal Sound Recording and Reproduction, Vol. 4 No. 5 (May 1954), pp. 131-138.
22-24: André Millard, "America On Record: A History of Recorded Sound" (*book*); Cambridge University Press, 1995, commencing page 147.
25-26: Scott Eyman, "The Speed of Sound - Hollywood and the Talking Revolution" (*book*), New York: Simon & Schuster 1997. From the point of view of the evolution of sound recording practice, the information is spread in a somewhat disconnected fashion; but relevant mentions of "The Jazz Singer" may be found on pages 12-15 and 135, "Old San Francisco" on pages 128-9, and Laurel & Hardy on p. 334.
27: (ed. Elisabeth Weis and John Belton): Film Sound Theory and Practice (*book*), Columbia University Press (1985): article "Sound Editing and Mixing" by Mary Ann Doane, pp. 57-60.
28: D. P. Loye and K. F. Morgan: "Sound Picture Recording and Reproducing Characteristics" (*paper*). Originally presented at the 1939 Spring Meeting of the S.M.P.E. at Hollywood, California; printed in the Journal of the Society of Motion Picture Engineers, June 1939, pp. 631 to 647.
29: Bela Balazs, "Theory of the Film: Sound" (*article*), in "Theory of the Film: Character and Growth of a New Art" (*book*); New York: Dover Publications, 1970. Reprinted in: (ed. Elisabeth Weis and John Belton): Film Sound Theory and Practice (*book*), Columbia University Press (1985), pp. 106 - 125.

Appendix 1. Preparing media for playback

With these appendixes, I must make it clear that I am giving my current favourite recommendations, not recommendations based on points of principle. It is almost certain they will change with time, and I advise you to keep abreast of new ideas. Thus I do not guarantee results, nor can I be held responsible for disasters.

Cleaning grooved media

There are several motives for cleaning a record - e. g. to improve its appearance for sale, prepare it for antistatic treatment, or seal it from atmospheric contamination - but I shall assume only one reason. That is to get the optimum power-bandwidth product from it. The aim is to remove as much foreign matter as possible without damaging the groove, so the stylus makes intimate contact with the groove.

I shall assume playback with a jewelled stylus in a pickup with relatively low effective tip mass, but possibly using quite a lot of downward pressure. Cleaning recommendations will not necessarily be correct for other methods, such as fibre needles or laser-beams. And I shall ignore side-effects like damage to the label, or the removal of "preservatives" or antistatic treatments which might previously have been applied.

I normally recommend that novices transfer each record twice, once before it is cleaned and once afterwards, so the maximum possible power-bandwidth product may be selected. After you have some experience, you will be able to say when cleaning is definitely a good idea and when it is not; but frankly there are so many cases in which cleaning might be a disaster that I recommend "before-and-after" transfers whenever you're not sure.

Washing disc records

Water is the best way of removing dirt, since most kinds of dirt dissolve in it. As far as I know, there is only one type of record which is directly attacked by water, and that is the "gelatine" disc dating from about 1935-1940 in the UK and later on the Continent. This served the same market as "nitrate discs," namely one-off direct-cut discs made by semi-professionals. It is not easy to describe the difference between a gelatine and a nitrate. If the disc has a brass ferrule round the centre-hole, that usually means a gelatine; but the reverse isn't always true. Those with a good sense of smell also tell me that you can smell a nitrate, but not a gelatine. A drop of water on the surface of a gelatine always makes it go tacky, so try this on an unmodulated bit first. (Unfortunately, I do *not* know how to wash a gelatine!)

For all other records, large supplies of distilled or demineralised water will be required. The action is to coat the surface with water which has reduced surface-tension, so the water makes good contact. Photographic wetting-agent or pure liquid detergent can be used, with the latter having the advantage of dissolving grease when it is encountered. (But do not use detergent with scented lemon additive or other potions!) It also helps if the water is at or slightly above blood-temperature, since this automatically means that the grease of human fingerprints is removed. But I must warn you against submitting cylinders to thermal shock. All cylinders, whether direct-cut or moulded, have a large coefficient of thermal expansion to enable them to be withdrawn from a mould. Sudden application of only five degrees of temperature change can split a cylinder. Allow

it to acclimatise first, perhaps by unloading it from its box in a microenvironment just above blood temperature, perhaps on a shelf above a convector-heater.

Risks of harm from water

Look out for obvious cases where water will cause damage. Besides the gelatines I have already mentioned, these fall into two classes:
1. The label. Labels on pressed discs will survive, because they are pressed into the disc at the same time as the grooves, and will not come off except by vigorous manual scraping. They will however lack their new glossy appearance after a wash. Lick-and-paste labels (e.g. copyright stamps, and labels on most nitrates) will inevitably soak off. If details are handwritten in ink, this may become illegible as well. You are advised to photostat them first.
2. The core of the record. Many types of records have playing surfaces unaffected by water, but layers underneath are vulnerable. Blue Amberol cylinders and some other makes have a Plaster-of-Paris core, which will be exposed if the plastic has been chipped or cracked. Columbia Laminate discs are made of three layers cemented together with kraft-paper; this may swell and cause incipient splittage if the water gets through the shellac. (Learn to recognise a Columbia laminate from its appearance; it is difficult to describe, but it is blacker and has irregularities in the order of half a centimetre in its surface flatness). Avoid washing any "unbreakable" record based on cardboard or paper, which you can tell by its lightness and the sound it makes when it is tapped. Edison Diamond discs are also of laminated construction. As the edge is "squared-off" rather than rounded, wear-and-tear inevitably results in leaks, and water can trigger delamination even though the core itself may not be affected.

I have heard it reported that many shellac discs can be harmed by prolonged immersion in water. The argument is that there are hygroscopic particles in the mixture, and water will cause them to swell, increasing the volume of crackle. I must say I have never noticed any such problems myself, but most British shellac records are very crackly anyway. Maybe it's our damp climate and the records have already been attacked. But it seems reasonable to prepare yourself so you can dry the record immediately after it has been washed (within seconds).

The actual washing process

Next, we must dislodge particulate matter from the surface so the water holds it in suspension until it is rinsed off. This is where we will scratch the record if we aren't careful. For disc records, I prefer to hold the disc horizontal beneath lukewarm flowing tap-water, gently brushing the surface with my fingers. The tips of the fingers are the most sensitive parts of the body, and it is easy to detect when you have a grain of abrasive material between your finger and the record. You can then stop the brushing action and concentrate on removing the grain. For shellac records a fingernail is harmless; but for vinyl or nitrate it will be necessary for the water to drain away where the grain is located, so that it may be found by eye; then it may be dislodged. (Personally, I use the corner of a blunt razorblade, precisely like the ones used for magnetic tape editing and having become bluntened by a hundred edits or more!) The idea is that if you can see the grain, any damage caused to the disc is very localised, and much preferable to the large area attacked by a fingernail. Declicking processes at a later stage will therefore cause less corruption to the wanted sound.

Many cylinders are made of materials with practically no tensile strength, so it is difficult to grip them and wash them in a way which does not risk breaking them. However, they resist compressive forces comparatively well. I find the best way is to hold them under the lukewarm flowing water gripped between the thumb and forefinger of both my hands at once. By alternating the compression from one hand to the other, I can rotate the cylinder in my hand without submitting it to any tensile stress, and I can attend immediately to any grit. The grooves are usually too shallow to need a brush. Many kinds of cylinders seem very prone to mould. This is a case where I would *always* advocate transferring the sound twice, because the infection sometimes puts its roots so deeply into the wax that removing it only makes matters worse.

Most lateral-cut discs have comparatively deep grooves, and plain washing and massage does not always remove smaller particles down in the grooves. The next step is therefore a brush. The brushes which come with the Keith Monks Record Cleaning Machine are ideal for this; the nylon bristles are of an optimum size, shape, and consistency for brushing *down* into the groove without causing damage. Messrs. K.A.B Electro-Acoustics (1210 East Seventh Street, Plainfield, NJ 07062) also make a suitable brush. There is a fine distinction to be made in the stiffness shape and dimensions of the bristles. To avoid the risk of damaging many types of grooves, do not use *any* old brush; but the majority of shellac 78s will withstand anything!

Nevertheless, for soft discs such as nitrates (particularly microgroove nitrates), any brush will cause damage when dust-particles abrade the material. At the British Library Sound Archive we have an ultrasonic cleaning tank with vibrates dirt out of the grooves, but this isn't as effective as brushing for vinyl or shellac. It consists of a perspex tank-within-a-tank, the inner tank being of semicircular cross-section containing the pure water and wetting-agent, and the outer tank being tap-water whose only function is to conduct the ultrasound to the inner tank. The disc is lowered vertically into the inner tank on a pencil (you can often keep the label dry this way), and the transducer switched on. If you look down into the inner tank when this happens, the water often changes colour immediately, so something is certainly being vibrated out of the grooves! The disc is turned on the pencil so its entire playing area is washed.

Nitrates made by manufacturers in the USA during the 1940s and 1950s frequently exude plasticiser which causes difficulties to present-day operators. I am told the solution is to wash the disc in light mineral oil, but I do not have practical experience of this.

For washing vinyl discs, the Keith Monks people recommend four parts of distilled water to one of pure industrial methylated spirit. The latter is difficult to obtain without an excise licence, and personally I have found no disadvantage in substituting isopropyl alcohol. The machine applies the mixture through the fibres of the aforementioned brush, which speeds the process. But do not apply this mixture to other formulations of disc; there is good evidence that some shellac records (and cylinders) contain chemicals which dissolve in alcohol.

A brush is usually the *only* way to remove mould from discs and cylinders. In the case of cylinders, we use an ordinary half- inch paint brush with the cylinder in its dry state. Cylinders are difficult to brush without submitting them to tensile stresses. We have an old phonograph mandrel to push into cylinders specifically to solve this difficulty.

Drying the record

The next bit is the most difficult - to get the record dry after washing without sludge being deposited back in the grooves. The Keith Monks machine sucks the liquid off the

disc by means of a vacuum pump, and "sponges" the grooves by a continuously-wound thread of cotton. Because the disc is wet on both sides at this point, it is preferable to use the double-turntable model, one for the A-sides and one for the B-sides, to prevent any freshly-dried sides getting wet again.

The only practicable way to dry a cylinder is to roll it over sheets of blotting-paper, or paper kitchen-towels of the kind with sufficient strength not to tear and leave fibres in the groove.

Finally, don't put your newly-cleaned records back into dirty sleeves!

Flattening warped discs and cylinders

It is probably better not to put any grooved media through a heating process if you can get round the difficulty somehow. On discs, lateral warpage may occur when the vertical warpage is cured. To put it in plain English, the pickup may go from side to side even though it no longer goes up and down. I do not know how this can be cured; but if you are stuck with a record which has previously been flattened this way, wow is minimised (not eliminated) when the disc is carefully centred and played with a parallel-tracker.

I therefore start by urging you to consider ways of flattening a warped record mechanically, without using heat. A centre-clamp on the turntable helps. Turntables of the same size as the disc concerned (e.g. ten-inch ones) are useful, because you can then pull raised sections of the record's rim down, and hold them down with adhesive tape. (Don't stress shellac discs beyond their breaking-point!) Alternatively, keep a handful of flat unwanted nitrate discs to hand, preferably the ones with steel bases rather than aluminium, and tape warped discs to these. (Steel discs will deflect less than aluminium under the stresses involved).

A disc-cutting lathe with a "vacuum turntable" is probably the only way to control some of the flexible discs of the 1930s with their inherent tendencies to assume surrealistic shapes. The vacuum needed for sucking down a nitrate had to be restricted in power to avoid damage to the other side, so you may have to install a more powerful vacuum system if you follow this plan.

Half-speed copying may also make a warped record playable when it wasn't before (section 5.2).

However, these temporary remedies do not *cure* the warpage, so wow is inevitable, and there remains a risk that warped shellac discs will crack when shelved with flat ones. Thus we may be forced to heat the disc and risk the geometrical distortions.

Most types of conventional shellac disc can be flattened by placing them on a sheet of plate glass and gently heating them. (At the British Library we use the fronts of old monochrome television sets, which had a heavy sheet of plate glass between the viewer and cathode-ray tube to minimise the causes - and effects - of implosions. For the oven, we use the Gallenkamp Hotbox oven Size 1, although this is too small for some large transcription discs). If you lack an oven, the job may be done under warm water, or even in the sun; but it is much more difficult to control the temperature.

There are two secrets to success:

1. The glass and the record must be scrupulously clean, or any dirt becomes pressed into the surface and cannot be removed.
2. The heat must be just enough to allow the shellac to sink flat under its own weight and *no more*.

Now some general comments. Some discs (e.g. laminates) will not sink under their own weight and will need downward pressure from another glass sheet. But never ignore

the basic principle of using the lowest temperature possible, because slightly too much will cause the surface to go grey and very hissy. So monitor the proceedings carefully. A temperature in the neighbourhood of 42 Celsius seems about right, but do not trust the oven's own thermostat, as this temperature is very critical (and varies from one make to another). With some ovens the difference in temperature between the bottom and top shelves can be fatal. The Gallenkamp ovens have a fan inside to ensure uniform temperature, but even this isn't the complete answer because it can take quite a long time for plate glass to warm up (half an hour or more). During this time the thermostat indicator shows the air has reached the desired temperature, and you may be tempted to overcook the discs if they have not yet sunk under their own weight. Perhaps the best method is to remove a disc as soon as it falls flat, irrespective of what any other discs in the batch are doing. Fortunately it seems impossible to cause any harm by opening the door and having a look! Experiment carefully with sacrificial discs of the same make and age if you can. An electronic temperature-probe with a display outside the oven is useful. You do not need a high degree of absolute accuracy, so long as you use the same probe all the time and build up a record of the figures you have used.

Some shellac discs have ridges surrounding the playing-surface to protect the grooves from surface abrasion. This process will press these ridges flat. On single-sided records, keep the grooves uppermost and this should not occur. But note that on double-sided records this process will successfully flatten the playing surface at the cost of permanent distortions where the ridges are. This means that you must put the side with the smaller inner ridge downwards, or the upper playing surface will be distorted.

If the disc is pretty battered, you may like to try heating it to a fractionally lower temperature and pressing it down on the glass yourself. For curing warpage this is perfectly satisfactory, but it allows the ingress of dirt. It is also the safest method when you lack accurate temperature control.

After the disc has sunk to the flat state, slide it out of the oven on its plate glass and allow it to cool naturally (if necessary, while you're heating the next one). Do not take the disc off the glass until it's absolutely rigid.

Vinyl discs should go through a similar procedure, but note that the material has inbuilt stresses dating back to when it was made. These will mean the disc must be forced flat by a second sheet of glass on top. Fortunately, vinyl is more flexible than shellac, so the extra pressure should not cause the disc to crack. Cleanliness is particularly important with microgroove records of course, and frankly I do not recommend this treatment unless the disc is so badly warped that it is unplayable any other way. Both the glass and the disc must be perfectly clean, by which I mean free from dust-particles; this is quite difficult to achieve. The temperature needs to be slightly higher for vinyl - around 48 degrees Celsius. It may be necessary to aid the flattening process with weights, but in my opinion a second sheet of quarter-inch plate-glass is heavy enough by itself.

The only exception to this is where the disc has acquired many "bends", e.g. it has tried to take up the shape of the back seat of a car in the sun. Since these bends will be abrupt and will contribute much low frequency noise when the disc is played, it will be necessary to decide how the disc will be reproduced. If you have access to half-speed turntable and a parallel-tracker, leaving some ripples will be preferable to forcing the disc to go flat when it doesn't want to.

In the author's experience, sandwiching the disc between two sheets of plate glass (this is the lowest practicable weight) still means that the inclines from the back seat of the car turn into large lateral distortions in the circularity of the grooves. Thus, additional weights to cure the vertical ripples will cause very substantial lateral distortions, and it will

be practically impossible to play the disc with a conventional pivoted tone-arm at any speed. Thus you will need a clear understanding of the properties of the machine you propose to use before taking irrevocable action.

The unrelieved stresses will often cause the disc to warp again when you are not looking. The only method I know of coping with this is to keep the disc at the same temperature for at least 24 hours continuously. So you will need an oven devoted solely to this!

This flattening process cannot work with 45rpm discs and others with a raised label-area. Fortunately, most 45rpm material exists on other media with a better power-bandwidth product. I have not tried it myself, but the only alternative is to use two pieces of plate glass with holes about 95mm diameter cut in them, so the groove areas are flattened without affecting the label.

For a warped cylinder (made of homogenous material rather than with a Plaster-of-Paris base), it is possible to put it on a phonograph mandrel and gently heat it with a hair-dryer while it is rotating. Do not overheat the cylinder; a temperature of about 100F or 35 Celsius should be the maximum. (You could even put the working phonograph and cylinder inside the oven to be sure of this). As the cylinder warms up, push it further onto the mandrel until it is substantially round. The difficulty is to get it off again without its splitting when contracting. Switch the hair-dryer or oven to "cold," and be prepared to nudge the cylinder towards the narrow end by a millimetre or two every few seconds.

If this seems risky, you are right; but you should see some of the recipes for straightening Plaster-of-Paris cylinders! (Hillandale News, December 1964, pages 96-97; April 1968, pages 235-237). I have no hesitation in restating the principle I started with, "play the record in its warped state if you possibly can." I have a phonograph modified so the *cylinder* is driven by a light belt near the pickup, rather than the mandrel. Thus constant linear speed is assured under the pickup, provided the cylinder fits onto the machine at all.

Badly Cracked or Scratched Disc Records

In section 4.12, I dealt with some of the problems of cracked or broken records which throw a pickup out of the groove. If you have more than a few such records, it may be worthwhile to dedicate a special turntable with two pickups to the problem. The geometry for their layout needs to be carefully thought out, because the idea is to have one pickup controlling the other.

At its most fundamental version, the turntable is given two pivoted pickup arms with their pivots about four inches apart parallel to the radial movement of the first pickup (which will be playing the wanted sounds). The second pickup is used to play a larger disc (we keep a number of "sacrificial" sixteen-inch discs for just such a purpose), which is placed under the first disc. For convenience, both cartridges should be the type mounted in shells which have been drilled to reduce their mass. Choose a large disc of approximately the same groove-pitch as the smaller disc, and rig a piece of copper wire from one shell to its neighbour such that it will both pull and push without chattering or bending ("Blu-tac" may be needed to stop any chattering). Increase the playing-weight of the first pickup cartridge by the traditional pencil-eraser, and as it plays the big disc it will prevent the smaller disc throwing its pickup out of the groove for a number of revolutions, before the remaining differences in groove-pitch cause it to jump.

Broken Records

Quite honestly, the best way of dealing with a broken record is to find another copy. And failing that, we at the British Library pass the job to a specialist freelance. But for those of you with time to spare and the urge to experiment, I summarise briefly John R. T. Davies' recommendations on the subject.

1. If possible, carry out the restoration soon after the breakage has occurred, so as to reduce the chances of inbuilt stresses distorting the bits.
2. Collect the bits in separate polythene bags to prevent the sharp edges abrading together, increasing the size of the crack when they are reassembled. Look everywhere for *all* the bits!
3. JRTD has a jig for assembling the bits, which comprises a circular "sub-turntable" about half an inch thick, tapped at regular intervals with screw-holes. The disc can be assembled and held in place with clamps holding down each bit, and sometimes this is sufficient to get a play. Sometimes, too, the naked eye can achieve an acceptable alignment of the bits; but a 40-diameter microscope is a useful asset.
4. If the breakage is more complex, so that simple clamps cannot hold the bits together, it is necessary to stick the record together, building it up over several days to allow the glue to dry each time. "Araldite" epoxy resin is used. To prevent the resin from forcing the pieces apart, a narrow channel is cut into the *middle* of the cross-section on each side of the crack, and the resin is injected into these channels.
5. If chipping or stress-relaxing has meant the grooves do not match accurately, a hot needle can be used to apply wax from the core of a "Chinagraph" pencil into the crack. The surface of the wax can be carefully cut away to leave it flat, and it is even possible to recut grooves to prevent groove-jumping. Sometimes one can (in effect) rebuild missing groove-walls this way, thus at least getting a playback, even though the original sound is of course lost. Conventional de-clicking techniques are used to clean up the resulting sound.
6. For cylinder records, I am indebted to Michael Khanchalian for another approach. He is a dentist, and uses dental tools and dental inlay wax for repairing breaks. He has even had some special dental wax made, coloured light brown or black, so you cannot see the join!
7. In general, it is not possible to restore a disc record if there are two or more sectors missing. One sector of a disc can be dealt with in the following manner. On the main turntable, place the sub-turntable tilted at a slight angle by the insertion of a couple of felt wedges opposite the missing sector. Adjust things so the missing sector is at the lowest point, and the two edges of the missing sector are at the same height. The pickup-arm pivot is raised a corresponding amount, and a carefully-aligned metal bar is installed on the deck in a vice arrangement to prevent the pickup arm dropping down into the gap. When correctly set up, this arrangement will play all the grooves there are, and leave silent passages where the missing sector was. Conventional editing techniques can then be used to fill the gaps as required.

Preparing Tape for Copying

In Chapter 6 I discussed incompatibilities of tape speeds, track-layouts, and equalisation; but I did not address the problem of fitting the spool onto the machine. There are three standard ways: the "cine-centre" spool, the "NAB-centre" spool, and the European turnbuckle. Most quarter-inch machines are provided with adaptors for the first two, but

turnbuckle centres may be found in the middle of pancakes of tape without flanges, and in single-sided "reels" with only one flange. Clearly you should acquire a handful of the necessary bits which to allow a turnbuckle to fit on a cine-centre drive. These accessories are no longer made, and if you cannot find a set, you may need to place the tape on (for example) a disc turntable, and pull the tape off onto an NAB reel on a nearby tape machine. Flangeless pancakes will need a sacrificial disc underneath them. You may also find pancakes with more tape than a NAB reel will hold. What you do about these depends whether you can cut the tape or not; but it isn't impossible to *play* the tape directly from a disc turntable. The idea is to take up most of the tape onto a NAB spool, and allow the remainder to pile up on the floor. This is much less hazardous than it sounds, so long as the tape machine is four or five feet above the floor so the tape falls naturally under its own weight, and you do *not* touch the pile. The tape can then be wound back onto the disc turntable at 33rpm through the guides of the tape machine without tangling itself. Creative use of matchsticks and gaffer-tape may be needed to fool the tape machine's motors and brakes.

Now to problems of the tape itself. The first is that "acetate" tape can be very brittle. It breaks cleanly, so you do not lose any sound because a simple splice is sufficient to get it back; but you should know about the matter in advance. If the tape is on a small reel with a small core, the back-tension of the tape-machine may break the tape as it nears its end. The tape should first be spooled onto a reel with a bigger core, perhaps with extra non-acetate tape to increase the diameter. "Acetate" and "tri-acetate" tape can be recognised by holding it up to the light. If diffuse light comes through, then the tape is acetate-based. Other formulations are completely opaque.

"Acetate" tape can also warp. Here the best solution is a tape reproducer with variable back-tension. Ours is a Studer B67 with an additional printed circuit board governing the spooling, but again creative use of string and weights hanging on the feed-roller can have the same effect. A Cedar azimuth-corrector (section 4.16) will improve a lot of timing-errors as a result of tape not being perfectly straight.

Tape can also get mouldy if stored in dampish conditions. During the early 1970s much professional tape was sold with a matt finish on the back, because this helped a neat pancake to form at high spooling speeds. This matt backing nourishes mould, which the author has found particularly damaging on Scotch 262 and Emitape 816, ironically top-of-the-range tapes of their time. Usually it is necessary to replace the tape box and often the reel; but the oxide does not usually become damaged. My solution is to keep an old tape deck (actually a Studer B62) with worn tension-arms between the reels and the rollers. The prime function of the arms is to keep the tape taut against the rollers irrespective of small variations in tension at spooling speeds; but they inevitably acquire box-shaped notches where the tape wears them away. If the arms are *not* replaced, they make ideal tools for scraping mould off, which can happen as you spool. The only precaution must be to wash the machine afterwards and vacuum the floor to stop other tapes being infected.

This last-mentioned process is also the most satisfactory way of cleaning tape which has been in floodwater, or has otherwise got sludge onto it. Both ferric and chromium-dioxide tapes are completely unaffected by water, and this technique of "dry cleaning" is perfectly valid. In the event of flood damage, experiments by the author have shown that flash-freezing the complete package doesn't affect the sound, and thus it may be possible to rescue documentation. I wouldn't use it on metal (iron-particle) tape though, both because we can foresee a failure-mechanism (rusting of the iron), and because such tapes must not be subject to the least geometrical distortion as the

individual tracks and the recorded wavelengths are so tiny. I regret I have not found a foolproof method of cleaning metal tape; in any case, fatal damage has probably happened by the time *you* get it.

Normally, the dry-cleaning process is more of a health-and-safety problem for the operator than for the tape itself. So I must remind you not to forget the views of your staff and your practices about such things as face-masks, bulk air filters, etc.

"Sticky-Tape Syndrome"

This has had a lot of publicity recently. It is a phenomenon where synthetic polyurethane binder (for binding the magnetic oxide to the backing) absorbs moisture from the atmosphere and becomes sticky. The symptoms are usually that the tape starts playing satisfactorily, but after a few seconds an invisible layer of binder builds up on the tape heads, the tape starts squealing in protest, and finally judders to a halt. The syndrome affects several makes of tape marketed between 1979 and 1983, including videotapes.

The treatment is to hold the tape at a temperature of 45 to 50 degrees Celsius to drive off the volatile components. It will then be possible to play the tape and make a copy of the sound shortly afterwards, but stickiness will eventually recur. Personally, I prefer to start by winding the tape from one reel directly onto another (without going through tape-guides or past tape-heads). This gives a slack and irregular wind to help the volatiles escape; but watch this process carefully. Sometimes the binder has oozed so much that it sticks the oxide onto the back of the next layer, and of course the oxide comes off and you lose the recording. This would have happened even if you had tried a straight playback. The only cure I know is to run from one reel to another very slowly and at extremely *high* tension. The layers then separate slowly and tangentially without damage to the oxide; but you will need a specially-modified tape deck with servo control of slowly-rotating spooling motors, and it may take a day or two for one reel to unwind. Do this in a warm room to allow some of the stickiness to dry immediately. We have a prototype machine (called "the Grandfather Clock"), which supplies warm air to dry a few feet of tape as it crawls from one reel to the other.

For less extreme cases, the National Library of Australia has researched the effectiveness of the drying method, and for quarter-inch tape, recommends it should stay at 45-50 degrees for between eight and twelve hours. As usual, soften any "thermal shock" to prevent rapid expansion and contraction as the temperature changes; I keep several sealed bottles of water in my oven as a "thermal sink." I would imagine that longer periods may be required to allow the volatiles to evaporate from wider tapes, or from tapes packed in cassette shells (such as Umatic videocassettes). For audio, the actual temperature is not critical, but high frequency losses become measurable (although not usually audible) at temperatures up to 80 degrees, at which temperature plastic reels warp. For video, use the minimum temperature you possibly can; the wavelengths of video are so short there is a real chance the picture will be destroyed altogether.

I am told a similar effect can also occur on audio tapes from the early 1960s, but I have no personal experience of these; nor can I say whether the cure is the same. Different workers have different ideas about the details of the process, some of which I will list now.

(1) Use a "plain oven", not a microwave oven. One with a circulating fan is best. (The British Library Sound Archive uses Gallenkamp Hotbox Size 1, the same as for flattening warped disc records). But it is remarkable what can be done with a hair-dryer and a shoebox!
(2) For nastier cases, it may be necessary to add another processing stage, namely spooling it gently past 3M's Type 610 Professional Wiping Fabric held between finger and thumb. This may have to be iterated several times until the fabric stays fairly clean through one pass.
(3) The dried tape should then be played (and the sound recovered) fairly immediately. It should continue to be playable for a month or so, but will eventually become sticky again.
(4) Once a batch of tape has been identified as having "sticky tape syndrome", resources should be allocated to getting the whole batch playable as soon as possible, because things only get worse the longer you wait.

Appendix 2: Aligning analogue tape reproducers

This section does not teach you how to make an analogue tape *recorder* work at its optimum, only a tape *reproducer*. And it does not say anything about freakish tapes which have been recorded upon wildly faulty machines. At present, correcting faulty tape *recordings* nearly always involves a subjective element. There is of course no reason why you should not do your best to simulate the original sound for the purposes of a service copy; but in anticipation that future operators will have access to better technology (if not expertise!), you should at least do an objective copy to the standards laid down by an international body, so it will always be possible to know what has happened in the meantime.

Chapter 6 mentioned that identical principles apply to audiocassettes, the "linear" soundtracks on videotapes and videocassettes, and magnetic film soundtracks. To save space, in this appendix I shall just use the word "tape" for all these media.

For the reasons I mentioned in sections 7.7 and 7.8, since the mid-1950s there have been two different standards for tape recording and reproduction, the "NAB" one (used in America, Japan, and by American companies operating elsewhere); and the "CCIR" one (later called "IEC") used in most of the "Old World." But two standards caused difficulties, particularly for organisations which covered both zones - for example, if the first machine bought by an Old World organisation happened to be an American one, and after that they had to match it! I mention this because it is virtually impossible to align a tape reproducer without an alignment tape of the correct standard (and tape-speed). So if you do not already have alignment tapes, acquiring the ones you need *must* be your first step. At present, I know no practical way of circumventing this.

Ideally you might buy engineering calibration tapes for two reasons:
(1) Playback frequency-response calibration (to the NAB or IEC standard), which I shall call "a frequency test tape" for short.
(2) Testing that the reproducing machine has sufficiently low speed variation, which I shall call "a wow-and-flutter test tape" for short.

Absolute Speed Test Tapes

Many pedantic engineers insist you must *also* have a tape to check the absolute (i.e. long-term) speed of the tape reproducer; but I disagree. Nearly all professional tape machines have a stroboscope which is sufficiently accurate anyway. You might think it better to use "high-tech" digital readouts from machines playing calibration tapes; and that certainly used to be true for manufacturers making hundreds of machines a week. But this is a case where a low-tech solution is just as good as a high-tech solution. It is trivially easy to make an absolute-speed test tape for yourself (and I still use some I made twenty years ago)!

The idea is to begin by bulk-erasing the tape. (I admit that in principle a bulk-eraser is the last thing you want in a sound archive, so it may be necessary to keep it locked away!) Such machines take a large amount of alternating current from the mains (in the order of a kilowatt for quarter-inch tape, rising to perhaps three kilowatts for two-inch tape). When switched on, they generate a powerful alternating magnetic field, usually emitted from the top surface of the device. Take your wristwatch off, and put the tape to be erased on top of the machine. You will hear a buzzing sound as the magnetic particles of the tape attract and repel each other; then lift the tape off the machine *slowly* (this is the difficult bit). As you go through the "B-H Curve" (Box 7.2 in section 7.2), you

will reach a stage in which you are going through the mid-point of the B-H curve, and quite suddenly the tape will be easier to lift. Yet it is just at this point that you have to move the tape *slowly*, to prevent some parts of the tape being permanently magnetised and other parts not, causing "thumping" noises.

When the tape is perfectly erased, simply unroll a length of it and lay it across a reasonably clean floor in a straight line (perhaps along a deserted corridor). It should be under tension similar to that of the reproducing machine you propose to use. (Something like 2 grams for cassette tape and 20 grams for quarter-inch, but modern tape has sufficient tensile strength along its length for this to be the least-significant of the various causes of error).

I shall deal with the easier format first, audiocassette. You do not need to unscrew the shell and extract the tape (unless you want to). Fewer tangles occur if you simply pull it out and wind it back afterwards, using a hexagonal pencil in the hub. Cassette tape is supposed to run at a speed of one-and-seven-eighths inches per second (4.7625 cm/sec), so one minute of tape should be 112.5 inches long (285.75cm). Touch the oxide side of the blank tape at right-angles with a magnetised razor-blade at suitable intervals. I prefer at least five consecutive minutes; and then get the tape back into the shell. Simply play the resulting cassette, listening for the thumps with a stopwatch in hand. Five minutes of thumps enables human reaction-times to average down to a sufficiently accurate level.

The equivalent open-reel tape will be much longer (at 15 inches per second, anyway)! But a location such as a football field at dawn, is perfectly satisfactory. Apply a leader-tape, hook it round one of the corner posts at ground level, and pull it straight across the field to the opposite corner-post to stop it blowing away in the wind. This will be a distance in the order of a hundred yards, which is roughly 3600 inches or 100m. A surveyor's measuring-tape will needed for something this long; but apply the magnetised razor-blade at intervals of (say) 900 inches (22.86m), which corresponds to 1 minute at 3.75 inches per second (9.525cm/sec) or half a minute at 7.5 inches per second (19.05 cm/sec). Use this tape in the same way as the audiocassette I've just described.

What Frequency Test Tapes Comprise

A frequency response test tape usually starts with a section of mid-frequency tone (usually 1kHz, being the middle frequency of suitable log/linear graph paper). Ideally this should correspond with the "zero decibels" mark on your test meter; but there are several approaches here (depending for example, on such things as Dolby alignment levels (sections 9.4 to 9.6) or the use of Peak Programme Meters for getting a reproducible result on actual programme material rather than steady frequencies). After the mid-frequency tone comes a high frequency for setting the azimuth of the playback head (section 7.6). After that will be a range of perhaps ten to twenty different frequencies, which should ideally all read the same; often the level of these might be different from the first tone. But the frequencies should all read the same on the meter.

Care of Test Tapes

Whatever you may think about their recorded content, engineering test tapes and discs are jewels, and should be treated accordingly. Apart from their sheer costs and the difficulties of using them without damage, they form the prime "standards" for your institution. They form much the same rôle for sound archives as (in Britain) the Weights and Measures Inspectorate uses for judging supermarkets! A tape machine should always

be fully de-gaussed before mounting one of these 'jewels', as any stray magnetic fields will compomise their performance.

However, tapes are relatively rugged, and do not need extreme climatic conditions for their storage. This is both their strength and their weakness. Ideally they should be available to all operators who might make "archive copies" at a moment's notice; yet if they are just slung in a drawer, they may become corrupted by non-engineers carrying (for example) magnetic microphones. On the other hand, putting them into a climatically-controlled vault may mean acclimatisation problems (let alone security hassles) when something goes wrong unexpectedly.

This writer's compromise is to make a number of "secondary" standard test tapes by the following process. First, record a number of suitable frequencies from an analogue oscillator onto a digital medium. This should have a greater power-bandwidth product than any analogue tape; but unless something is very faulty, the analogue-to-digital converter for your digital master will normally have satisfactory performance up to 20kHz, the maximum generally available from a primary test tape.

Make your digital master with each of the frequencies on your primary standard correctly (different for a frequency test tape and a wow-and-flutter test tape). That is, the frequencies should be the same as those provided upon your newly-purchased primary frequency test tape, and exactly 3180Hz on your primary wow-and-flutter tape.

You may sometimes find your "secondary" test-tape(s) may usefully be accompanied by "tertiary" one(s), of much shorter duration to provide a quick check at the beginning of each day. At the British Library Sound Archive, our secondary tapes also carry a short announcement *saying* they are secondary, and they end with a long run of the highest frequency, to allow the quick azimuth check described in section 7.6, and the subsequent screwdriver operations if a fault shows up. But the latter may be unnecessary if you *always* do the quick azimuth check by ear, *each time* you load a new tape on your reproducer to be digitised.

How To Manufacture "secondary" and "tertiary" frequency-test tapes

Check that this digital master reproduces all the frequencies to the same volume on the same analogue meter used for measuring the performance of the analogue reproducer for the frequency test. In the case of the 3180Hz which is the standard for the wow-and-flutter test, check it on a wow-and-flutter meter; it should easily be less than 0.05% weighted.

Having generated the "master" for the test tapes you wish to make, you should then take the primary tape to an analogue tape *recorder* with the analogue meter plugged to its *playback* output. The recorder should ideally have full-track record and playback heads, so the whole width of the tape is recorded with the same remanent induction. (This eliminates a source of variances, which I briefly touched upon at the start of section 7.11). But if you only have a stereo machine, most of the advantages of secondary and tertiary versions can be maintained at the cost of having to do twice as many adjustments.

At this point I shall interrupt the topic of making secondary or tertiary frequency test tapes. Ideally we should align our *recorder* to do the most accurate *reproduction* it can, which in analogue days was the main reason for *all* frequency-test tapes.

Different tape reproducers will have different user-adjustable amplifier controls, usually "presets" hidden under a flap for adjustment by a screwdriver. The first step is to identify the controls for "Replay" or "Playback", and not touch the ones for "Record"

until you've got the "Replay" matching your "primary standard". The next step is to find the ones for the correct tape speed, and not touch any of the others! If your organisation has a policy about reproducing tapes to a standard to make them all sound equally loud (which is normal for broadcasters), this will be the main reason for a "playback level" (or "playback gain") control. But as long as this doesn't corrupt the hiss or the distortion characteristics of the playback amplifier, or of reciprocal noise reduction systems downstream (section 9.4 onwards), your organisation may prefer to do such adjustments somewhere else (for example at a mixing console).

Nearly all reproducers have an adjustment for the extreme high-frequency response, using a pot labelled "HF" or "Treble". However, before you touch this, there might also be pots for "Mid" or "Bass", as well as a pot for "Gain" or "Level." Here I cannot give you a recipe-book approach, because sometimes the mid-frequency turnover needed for the appropriate "NAB" or "IEC" characteristic may be adjusted directly by just one of these controls, and on other occasions you may have to juggle two separate controls to get the correct turnover. The manufacturer's manual (for professional machines) or the Service Manual (for amateur machines) may be studied to find which control does what; but in the absence of this information, you may have to enter a period of serious experimenting with your primary test tape to get the response between 1kHz and 6kHz absolutely correct. This is much more important than any of the other adjustments, since it radically affects the response where the ear is most sensitive, and errors are much more conspicuous (even on poor loudspeakers). All the frequencies from 1kHz to 6kHz should be reproduced correctly, certainly within half a decibel, and ideally within a quarter of a decibel.

When you have got this right, *then* you may do fine adjustments to get the extreme treble or the extreme bass as accurate as possible. Here the results should ideally be within a couple of decibels for "service copies" and half a decibel for "archive copies."

Having done that, your next step will be to copy the digital tones from your digital medium onto virgin or "sacrificial" tape, which will become your secondary or tertiary frequency-test tapes. I advise you briefly to read section 7.3 to learn the qualitative principles behind A.C. Bias, and then I will leave it to you to adjust the *recording amplifier* controls to get the *reproduction* to mimic what your primary test tape did. (The exact levels of hiss and harmonic distortion aren't significant in this context; it is "simply" a matter of getting the frequency-response as "straight" as possible)

Since you are taking time in your battle to get acceptable performance, you might as well make several frequency test tapes while you are at it. At the British Library, we make ours in batches of ten. We keep at least one copy in each of our operational areas, and also make them available to our off-site contractors.

Conclusion

I conclude by reminding you about the *reason* for all this tinkering. If you propose to digitise analogue audio tapes, Chapter 6 tells you of some of the *compromises* made by the original recording engineers to minimise problems. At present, we do not have digital processes to correct these problems; yet it is absolutely essential that future archivists know exactly what we have done when reproducing the original tape.

I accept it would be very unfriendly to listeners, to force them to listen to a number of line-up-tones before the subject matter of their choice; but as long as you can correctly say that you have played the tapes to a certain characteristic, using a machine with less than 0.05% speed errors, then you can simply add a few characters to the document (and

its catalogue-entry), so future engineers know precisely what you did when you made the digitised substitute.

DISCLAIMER

The statements and comments contained in this manual represent the opinions of the author, which are not necessarily those of the British Library. The British Library shall not be liable for any losses or damages (including without limitation consequential loss or damage) whatsoever from the use, or reliance on the contribution, of the advice in this manual. Any links to other websites or organisations does not constitute an endorsement or an approval by the British Library of any products, services, policies or opinions of the organisation or individual.