DIGITAL AUDIO BIT-RATE REDUCTION FOR BROADCASTING

FRANCIS RUMSEY

University of Surrey
Guildford, GU2 5XH, U.K.

The bit rate of a single channel of high-quality digital audio may now be reduced by up to 90% using so-called 'perceptual coding systems' with minimal effects on sound quality. Such systems form the basis of proposals for digital audio broadcasting (DAB), this being the subject of the Eureka 147 programme, and they offer the possibility for extending the storage times available from digital recording products. Perceptual coders rely for their success on studies of psycho-acoustic masking, which suggest that a considerable amount of information in most audio signals is redundant as far as the hearing mechanism is concerned.

INTRODUCTION

The data rate required for a single channel of high-quality digital audio is considerable, being in the region of 750kbit/s before considering any overhead for coding. Significant advantages may be gained if this data rate can be reduced without impairing sound quality, especially in the field of digital audio broadcasting (DAB) where data-rate reduction is the key to achieving efficient usage of the radio-frequency bandwidth available. There are also potential applications for bit-rate reduced audio in recording, where a lower data-rate per channel may allow a longer recording time or a greater number of channels on a disk, for example, and in broadcast contribution and distribution circuits where greater use may be made of the ISDN (Integrated Services Digital Network) and other digital telecommunications services.

Recent research and development, both by independent companies and by those involved with the Eureka 147 Digital Audio Broadcasting project, has resulted in a number of systems which offer major reductions in the bit-rate required for high-quality audio, arguably with transparent handling of the audio signal, and with data rates usually between 64 and 192kbit/s per channel depending on the trade-off between data rate and perceived quality. This contrasts with earlier bit-rate reduction systems such as NICAM 728, now used for terrestrial stereo television broadcasts in the UK, Spain, Scandinavia and elsewhere, which do not achieve such a high degree of compression.

In the following paper an overview will be presented of the principles and applications of modern high-quality audio data-rate reduction in broadcasting. The detailed principles of individual systems will not be covered here, since they are to be the subject of subsequent papers.

1. RECENT HISTORY

Useful reductions in audio data rate, with minimal effects on sound quality, have been achieved in the 1980s by systems such as NICAM (a BBC-designed system) and Adaptive Delta Modulation (developed by Dolby Laboratories). NICAM (Near-Instantaneously Companded Audio Multiplex) involves the use of a sliding 'window' on blocks of 32 samples, which selects 10 bits for transmission out of an original 14 bits (linearly sampled). The maximum signal level within the block is used to determine the gain range used, and at low levels the 10 LSBs of each sample are transmitted whilst at higher levels the window slides upwards and LSBs falling outside the window are discarded (Figure 1). (This is the principle, although the situation is modified slightly by the adoption of two's-complement numbering, since the MSB represents the sign in this case.) The discarding of LSBs means that quantising noise is increased for high-level audio signals, but the system relies for its success on the knowledge that high-level signals will to some extent mask low-level noise (although this depends on the spectral content of the signal in relation to that of the noise). HF pre-emphasis is used before encoding, and de-emphasis after decoding, to reduce the audible effects of this quantising noise. Scale-factor information is transmitted alongside compressed audio data to allow for the restoration of samples to their correct ranges on decoding.

The NICAM 3 system [1] developed in the early

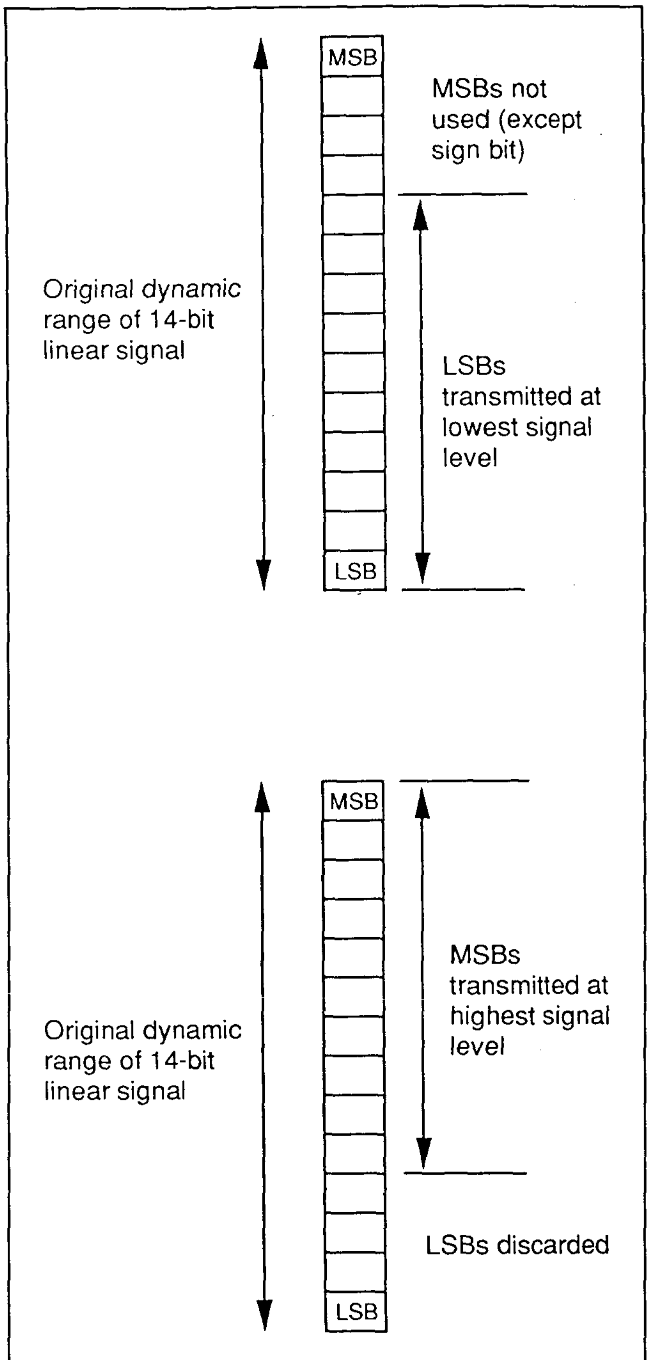

Figure 1 NICAM gain ranging (conceptual), modified in practice by 2's complement

eighties allowed for the multiplexing of six audio channels over a single 2048kbit/s digital telecommunications link, implying a data rate of around 350kbit/s per channel. Variations on this system have been employed in microwave link equipment for stereo outside-broadcast signals, and a form of NICAM operating at 676kbit/s is also used in the BBC's Dual-channel Sound-in-Syncs (DSIS) equipment used for the distribution of stereo sound with television pictures. NICAM 728 [2], operating at a total data rate of 728kbit/s, is intended for radio-frequency transmission of the sound for stereo television programmes, and uses a DQPSK modulation method.

Adaptive delta modulation [3], developed by Dolby, involves variable and fixed analogue pre-emphasis, as well as a delta-modulation encoder whose step-size is varied according to the nature of the audio signal. Again, the increased quantising noise which results is masked to some extent by high level audio signals, and the pre-emphasis is controlled to minimise the audible annoyance. Similar output data rates result as for NICAM, and the sound quality differences between the two systems are minimal. ADM has been adopted commercially in a number of areas, one example being in the Australian DBS (direct broadcast satellite) network.

These two systems are examples of modest bit-rate reduction using broad-band companding, the percentage saving of NICAM, for example, being only around 20%. (Original data rate per channel = 32 000 x 14 = 448kbit/s; compressed data rate = approx 350kbit/s). The latest systems allow the bit-rate to be reduced by a factor of between four and eleven, depending on the desired trade-off with sound quality.

2. PRINCIPLES OF RECENT BIT-RATE REDUCTION SYSTEMS

2.1 Masking models

In order to achieve the degree of data compression indicated above, recent systems have been based on a detailed study of psycho-acoustic masking. Such systems (often termed 'perceptual coders') rely on a model of the ear's masking threshold to determine which components of the sound are redundant' and therefore need not be transmitted, since they cannot be heard. The work of Zwicker [4] has shown that if the audio-frequency spectrum is split into 24 bands, each roughly corresponding to the critical bandwidth of hearing (that is the approximate analysis bandwidth of the first-level aural mechanism) then a masking model may be built in the form of a threshold-of-hearing curve which

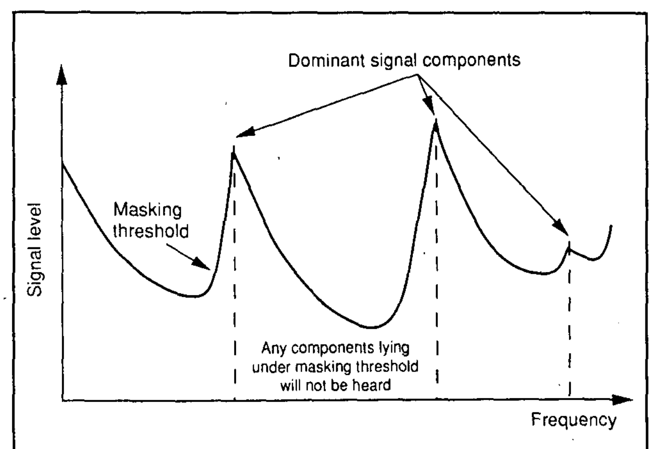

Figure 2 The threshold of hearing is modified in the presence of audio signal components (simple example shown)

depends on the signal being auditioned (Figure 2). It should be noted that Moore and Glasberg [5] have suggested somewhat narrower analysis bands than Zwicker.

It has been appreciated for many years that some sounds are masked in the presence of others, and even the non-acoustician will have experienced the difficulty of hearing low-level conversation in the presence of a nearby noise source. The threshold of hearing is raised around the frequency of a dominant signal component (the masking effect being more extended at frequencies above the dominant component than below it) and signals whose energy lies below this threshold are not normally perceived. There is an attempt in some systems to take advantage of the principle of 'temporal masking' which arises due to one sound's ability to mask another which occurs a very short time after it or before it. Pre-masking is less obvious than post-masking, and the effect is restricted to delays of less than 50ms before and after the signal in question.

Johnston [6] has given estimates for so-called 'perceptual entropy', suggesting values for the information content of audio signals based on a knowledge of the masking phenomena observed in the human hearing mechanism. Theoretically, the quantisation accuracy of an audio signal can be adjusted so as only to code samples accurately enough to fit the model of what can be perceived - there is no point in coding more accurately than is necessary. Johnston has proposed estimated minimum bit-rates required for totally transparent coding, and these are to some extent the goal of recent bit-rate reduction systems.

In most of the recent audio data compression systems the signal is split, either by filtering or mathematical transformation, into a number of narrow bands and then re-quantised to varying accuracy depending on the level of the signal in that band, based on a knowledge of which signals will mask others. The quantising noise is adjusted in each band (by varying the resolution of quantisation to an average of 2-3 bits per sample within each band) so that the noise lies just below the masking threshold for that band and thus will not be heard. A block diagram of the analysis and re-quantising operation is shown in Figure 3.

In passing it should be noted that the sound of the material which is *rejected* during data compression (that is, that which lies under the masking threshold) is most interesting to listen to, since one would expect that 85% of the original signal would sound rather more significant than it in fact does. The redundant material has a sound very much like pumping broad-band noise, following the rhythmic pattern of the original audio signal, but is really fairly inconsequential when auditioned in comparison to the transmitted 15% or so which appears to be all that is required for transparent coding.

2.2 Transform coding and sub-band coding

Two principal methods of splitting the signal into narrow bands are employed in perceptual coding, known respectively as 'transform coding' and 'sub-band coding'. There are also variations on and combinations of these two techniques. In transform coding, input audio data is subjected to a mathematical transformation during which a given number of time-domain input samples produce a number of frequency-domain output values, the amplitudes of which represent the relative strengths of different frequency components in the input signal over the duration of the transform window. Dolby's Adaptive Transform Coding system, for example [7], (not a Eureka contender) uses a 512-sample window and a modified discrete cosine transform (MDCT) to produce frequency-domain components which are then coded according to the relative importance and likely masking effects of each on the others.

In sub-band coding a digital filter network is employed to split the signal into narrow bands. The MUSICAM system, for example, splits the signal into 32 bands using polyphase pseudo-QMF filters which significantly reduce the number of

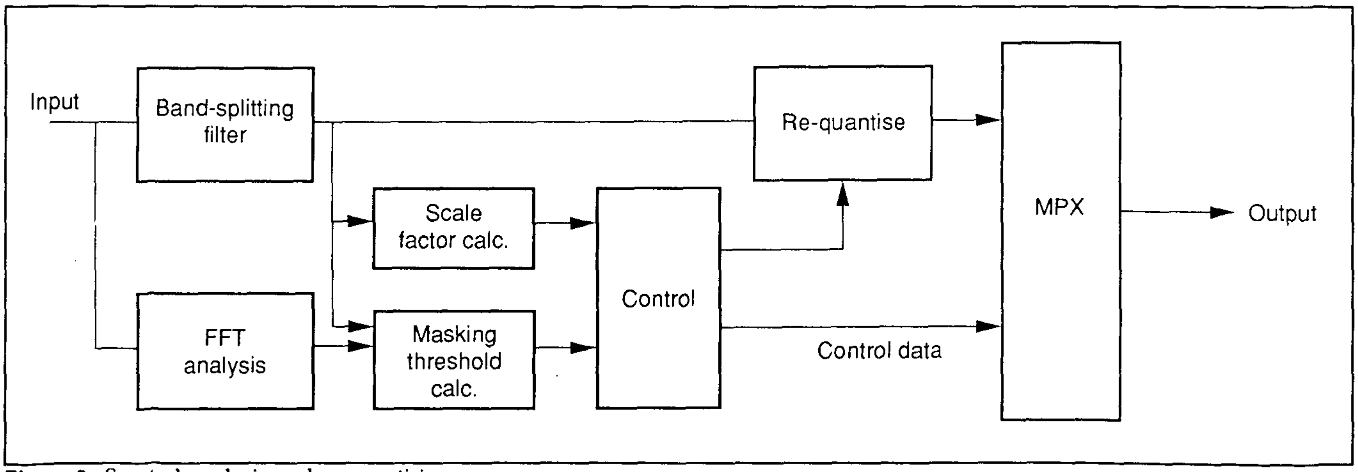

Figure 3 Spectral analysis and re-quantising process

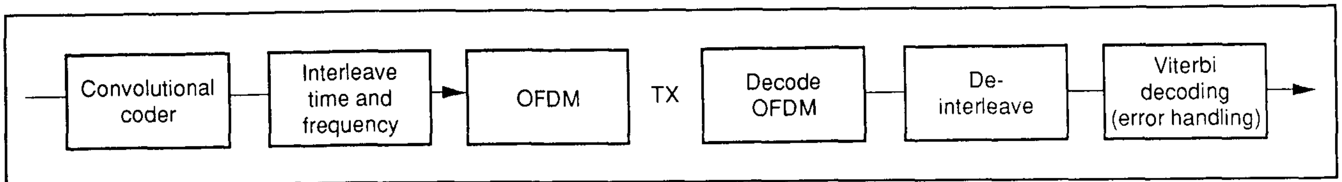

Figure 4 COFDM modulation and reception process

computations required when compared with standard QMF filtering, thus speeding up the coding process. Each of the bands in this case is of equal width (750Hz), although the psycho-acoustic model is based on aural critical bands which relate more closely to one-third octaves in the mid-frequency range.

In some systems the side-chain shown in Figure 3 which calculates the psycho-acoustic weighting parameters involves a fast-fourier transform (FFT), even though the band-splitting is performed using sub-band filtering. Most of the proposed systems introduce a coding delay of around 40ms, although the non-Eureka apt-X100 system [8] offers a significantly shorter delay of around 4ms, due to the use of only four sub-bands, in conjunction with processes of ADPCM, linear prediction and backward-adaptive quantisation which have their origins in speech-coding systems. By analysing the signal in the time domain this system looks for repetitive components which signify spectrally pure signals, and these are attenuated prior to quantisation since they tend to show up modulation noise more than broad-band random signals.

'Dynamic bit allocation' is used in audio data compression systems to manage the allocation of the available 'bit budget' to different frequency bands, depending on the prevailing psycho-acoustic analysis. The output data rate of broadcast systems is expected to be constant, and thus high coding resolution in one part of the spectrum will be traded off against lower resolution in other parts.

2.3 Data rate versus sound quality

It is possible to code a single full-bandwidth audio signal at a rate as low as 64kbit/s. Nonetheless, at this data rate some audible side-effects of coding are noticeable, although the sound quality is by no means poor. A rate of 128kbit/s is considered more appropriate for transparent coding, and 192kbit/s is considered to offer a useful margin over the minimum requirement for transparent coding.

Subjective tests have been performed on codecs to determine their relative merits at different data rates [9], as well as the effects of repeated coding and decoding. Critical listeners have suggested that the industry should be very careful about how quickly it embraces audio data compression in certain applications. Gerzon, for example [10], has pointed out that such systems measure much more badly than they sound, indicating that although audible transparency may exist, *measured* transparency certainly does not. At the same time, he points out that there are strong similarities between digital data compression systems and the analogue noise reduction systems which the industry has used for many years.

3. DIGITAL AUDIO BROADCASTING

Perceptual coding systems such as those outlined above form the basis of sound coding for future digital audio broadcasting systems. The Eureka 147 project aims to establish a worldwide standard for digital audio broadcasting, and progress towards this end is well advanced.

3.1 Transmission of compressed audio

The output of a perceptual coder is a collection of low-bit-rate samples representing frequency components of the input signal, and Eureka has chosen a modulation method devised by the French CCETT which will be used to transmit the compressed signal either from satellite or terrestrial stations. Coded Orthogonal Frequency-Division Multiplexing (COFDM) is a technique which involves the modulation of a number of input channels, representing individual radio services - intended to be between 12 and 24 stereo channels in the final version - onto a very large number of closely-spaced carriers (Figure 4). The input signals are subjected to an inverse Fourier transformation to allow the matrixing of components in both time and frequency domains in a form of two-dimensional pattern, with carriers overlapped in such a way as not to result in mutual interference. Interleaving and convolutional error correction is introduced in both time and frequency domains to make the transmission resilient to received errors. On reception the data is subjected to Fourier transformation to restore the information to the correct domain, then it is de-interleaved and errors are corrected or concealed.

A significant feature of COFDM is its ability to ignore multipath reflections, which have been the bane of FM transmissions (Figure 5). Reflections from buildings and suchlike produce delayed arrivals of the RF signal at the receiver and can result in comb-filtering of the RF spectrum and fading. Synchronisation patterns are included in the digital signal such that the receiver may

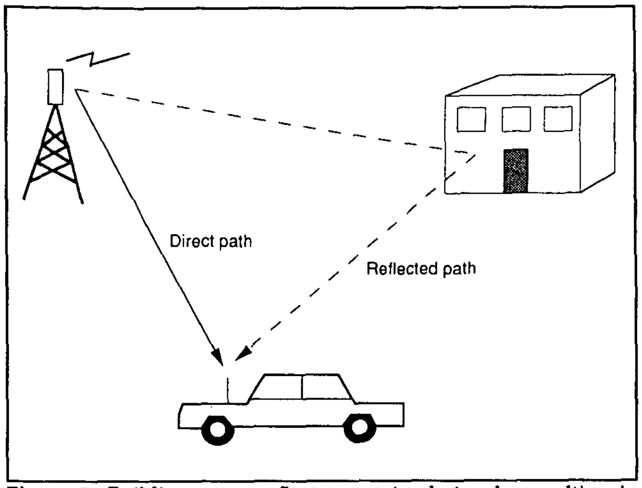

Figure 5 Buildings may reflect tranmsitted signals, resulting in delayed secondary signals

analyse signals received over quite a wide time window in order to extract information for each radio station. Reflected information is not detrimental, in fact the system thrives on multipath and co-channel information from adjacent transmitters carrying the same programme, since this simply provides additional redundancy.

3.2 RF spectrum planning

A number of possibilities exist for the assignment of digital radio channels to RF spectrum space. Between 16 and 24 stereo channels can be carried using COFDM in a 7 or 8MHz television channel, and this has been tested in France. It is possible, though, that 7MHz is too wide a bandwidth for many applications since it may not engender efficient use of band space when only small spaces are available (such as in Band II, currently used for terrestrial FM radio). 12 channels would be more sensibly carried over a 3.5MHz bandwidth for this reason. The question of RF allocation is to be discussed at the World Administrative Radio Conference in 1992 (WARC 92).

Proposals exist for transmitting DAB from satellites at either 1.5 or 2.5GHz, although many broadcasters would prefer the 1.5GHz frequency whilst the British Department of Trade and Industry would prefer 2.5GHz. The higher of these two frequencies would require some four times the transmitter power and four times the satellite cost, as well as increased usage of four 'gap filler' transmitters for RF-shadowed regions, in order to provide a similar level of service.

For terrestrial transmission, it has been suggested that the VHF frequency range is more appropriate than UHF, for ease of reception in cars and for ease of coverage of the country without complex receiver aerials and multiple transmitters. BMW, the motor manufacturer, has suggested that a frequency of 250MHz is the maximum which could be accepted for satisfactory reception in a vehicle travelling at 150km/h.

In Germany, space has been allocated at 230MHz for DAB, a frequency which is also sometimes used as a TV channel. The UK is looking for space in VHF Band III, and further work is needed before the clearance of space in Band II can be contemplated. An important factor is that maximum commonality of receiver design should be achieved, since although it is possible to construct receivers which will operate on a wide range of bands, these would clearly be expensive.

DAB transmissions are much more tolerant of co-channel interference, with protection levels of only some 10dB required for satisfactory reception (i.e. provided that signals received from transmitter 2 are at least 10dB below those received from transmitter 1, no interference will result), and this compares very well with the 45dB required currently for FM systems. Indeed if two DAB transmitters are carrying the same collection of programmes, very little problem with adjacent transmission areas would arise thanks to the multipath signal handling of COFDM.

In some areas it will be necessary to mount small 'gap filler' transmitters to cover RF-shadowed regions such as in built-up or hilly areas, and the number of these would depend on the frequency chosen and the decoder design. Tests in France have shown that the audio quality picked up in a moving vehicle receiving a COFDM DAB transmission is greatly improved when compared with the same vehicle's reception of conventional FM. Transmitter power can be reduced over that used for FM, with suggestions of needing only 1-2 watts, and fading will become a thing of the past. At last there is the possibility of receiving clean near-CD quality digital audio in the car, without the interference and fading effects experienced currently.

4. RECORDING APPLICATIONS

Data compression may be a solution to the problem of storing a larger number of channels or a longer programme time on magnetic and optical disks. Currently the capacity of optical disks is around 650Mbyte, and this is only satisfactory for around one hour of stereo in linear PCM format. Multi-channel removable disk systems could be developed using a proprietary data compression system based on perceptual coding, but the implications of such a development need careful analysis. In post-production signals may be copied many times, may be subjected to gain changes, EQ and speed or pitch changing, all of which may alter the psycho-acoustic masking characteristics of the signal. Furthermore, perceptual coding and decoding involves a short delay which could prove problematical where instant response from a recording system is required, such as during 'rock-and-roll' editing. The delay would also be difficult to accommodate in mixing and overdubbing operations.

Storage media with a considerably greater recording density than is currently available are on the horizon, and higher recording density may be preferable to data compression as a means of achieving greater storage capacity for original recording purposes, provided that access times to such media are sufficiently fast.

5. CONCLUSION

Perceptual coding allows the data rate required for high-quality digital audio transmission and recording to be reduced very considerably, with near-transparent audible results. The degree of transparency depends on the accuracy of the psycho-acoustic model and the target data rate. Its principal application lies in sound broadcasting, where it offers the possibility for a large number of stereo channels in a limited RF band, and applications may also exist in recording and post-production, provided that the sound-quality implications are carefully assessed.

REFERENCES

[1] Caine, R. et al: "NICAM 3: near-instantaneously companded digital transmission system for high-quality sound programmes" The Radio and Electronics Engineer, 51.10, pp 519-530, October 1980

[2] Rumsey, F.: "Stereo Sound for Television" Focal Press, London and Boston. 1989

[3] Todd, C., and Gundry, K.: "A digital audio system for broadcast and prerecorded media" presented at 75th AES Convention, Paris, 1984 preprint 2071

[4] Zwicker, E.: "Subdivision of the audible frequency range into critical bands" Journal of the Acoustical Society of America, 33.2, p. 248, February 1961

[5] Moore, B., and Glasberg, B.: "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns" Hearing Research, 1987, 28, pp 209-225

[6] Johnston, J.: "Estimation of perceptual entropy using noise masking criteria" ICASSP 1988, pp 2524-2527

[7] Davidson, G. et al: "Low-complexity transform coder for satellite link applications" presented at 89th AES Convention, Los Angeles, 1990 preprint 2966 (F-I-6)

[8] Smyth, S.: "apt-X100 sub-band ADPCM digital audio data compression" presented at National Association of Broadcasters Convention, 1989

[9] Gilchrist, N.: "Subjective tests on low bit-rate codecs" BBC Research Dept. Report RD 1990/16

[10] Gerzon, M.: "The gentle art of digital squashing" Studio Sound, pp 68-76, May 1990