REPORT 953-2*

DIGITAL CODING FOR THE EMISSION OF HIGH-QUALITY SOUND SIGNALS IN SATELLITE BROADCASTING (15 kHz NOMINAL BANDWIDTH)

(Question 2/10 and 11, Study Programmes 51C/10, 2F/10 and 11)

(1982-1986-1990)

1. Introduction

Technological developments presently allow the emission of digitally encoded audio signals to the public to be envisaged. It would be possible in particular to introduce this technique in the broadcasting-satellite service. This will require the prior definition of characteristics concerning coding, multiplexing and modulation of audio signals. This Report is concerned only with digital sound coding. Multiplexing, bit-error protection strategy and modulation aspects are discussed in Reports 954 and 632 respectively. Information dealing with methods of error correction or concealment can be found in Report 1073.

Study Programme 51C/10 considers that the introduction of digital techniques for emission should allow an improvement in the quality of the signals transmitted. Recommendation 651 deals with digital PCM coding. Results of objective and subjective tests are given in Annex II of the present Report.

Other coding methods leading to a substantial bit rate reduction and at the same time maintaining very high quality are described in Report 1199.

In addition to the improvement of quality, the following should be considered:

- the compromises between quality objectives and bit rate may be different for sound services which may have various quality requirements and planning constraints; they may also vary according to the requirements of individual countries;
- there are clear advantages for broadcasters, receiver manufacturers and the public in using a single standard for each application.

2. Digital PCM coding

2.1 *Linear coding*

The audio signal is encoded into digital form by a high precision analogue-to-digital converter after being passed through an appropriate anti-aliasing filter and possibly through a pre-emphasis network. The uniform coding resulting from this process leads to a linear representation with a minimum of 14 bits per sample. In that case, a 2's complement coding is used for each sample. A possible scale factor can possibly be associated with blocks of successive samples.

2.2 *Floating point or near-instantaneous coding*

A 16 bit linearly encoded sound signal can be transmitted on 14 bits per sample using a floating point coding system where the scale factor is based on 64 consecutive samples. Pre-emphasis may or may not be used.

When a lower bit rate per audio channel is required, the near-instantaneous companding allows for a reduction in the number of bits per sample from 14 bits to 10 bits. This companding is applied on blocks of 32 successive samples with a complementary scale factor on five ranges. A 2's complement representation is still used for each sample. In that case, the use of pre-emphasis is recommended to reduce the programme modulated noise.

---

* This Report should be brought to the attention of the CMTT.

2.3 *Differential coding*

Differential coding is another technique to achieve a lower bit rate per audio channel by making use of the fact that, in most programme material, correlation between two consecutive samples is high. In this case, the difference between successive samples will generally be smaller than the actual sample values. By coding the differential signal, rather than the original samples, a reduction of bit rate can be achieved for a given requirement of signal-to-noise ratio.

Difference samples can be coded instantaneously, one at a time, as in a linear coder, or by blocks of successive samples (with complementary scale factors) as in near-instantaneous companding. Such a marriage of differential coding with near-instantaneous companding has been proposed [CCIR, 1986-90] to achieve a reduction in the number of bits per sample from 15 bits to 8 bits. The use of pre-emphasis is recommended to reduce the programme modulated noise.

2.4 *PCM emphasis*

The quality of the transmission system, in particular the noise and distortion characteristics, depends largely on the signal statistics and the coding law used. Also, the audible noise detection response of the receiving system, including the human ear, can be considered to be non-uniform. Taking these factors into account, the characteristic could be improved by adopting appropriate pre-emphasis prior to the coding process and the corresponding de-emphasis at the receiver following the decoding process. For PCM systems two such emphasis methods have been exhaustively studied and evaluated. Annex I contains more detailed information on the two systems, i.e., one based on CCITT Recommendation J.17, the other based on 50/15 µs.

3. Adaptive delta modulation (ADM)

3.1 *Background*

The adaptive delta modulation (ADM) system is based on delta modulation (the decoder is mathemetically defined in § 3.2). The audio is encoded into digital form by a simple delta modulator after being modified by two processes:

- the audio is passed through a variable pre-emphasis network which alters the audio spectrum; and
- the audio is compressed in level, based on its slope. The compression is "infinite" over a 48 dB input signal level; that is, the audio is compressed to the same level for conversion to digital form. Optimum loading of the digital channel is achieved for most audio signals.

The digital bit stream representing the encoded audio signal is transmitted to the receiver, along with two very low data rate bit streams containing control signal information so that the decoder can precisely reverse the processing performed by the encoder. The decoder recovers audio by integrating the audio bit stream; performing a dynamic range expansion based on the compression (slope) control signal; and performs spectral de-emphasis based on information in the emphasis control signal.

The following points may be noted concerning the use of delta modulation:

(a) the bit rate of delta modulation may be reduced with a minor effect on audio signal quality (S/N degrades by 9 dB for a halving of bit rate);

(b) reproduced errors are tolerable;

(c) a digital-to-analogue converter for delta modulation is very simple, requiring no precision components;

(d) the de-emphasis provides adequate output filtering and sharp cut-off low pass filters are not required.

Items (c) and (d) affect the required decoding circuitry.

3.2 *ADM decoder definition (Fig. 1)*

3.2.1 *Audio decoder*

The audio decoder consists of a leaky integrator fed with pulses derived from the audio data bit-stream so that data 1's and 0's at a rate in the order of 250 kbit/s cause the output to move positively or negatively by equal steps. The size of the pulses is linearly proportional to an applied control signal over a range of about 50 dB, and the leaky integrator has a frequency response defined by:

$$(1 + sT_0)^{-1}$$

where $T_0 = 0.5$ ms.

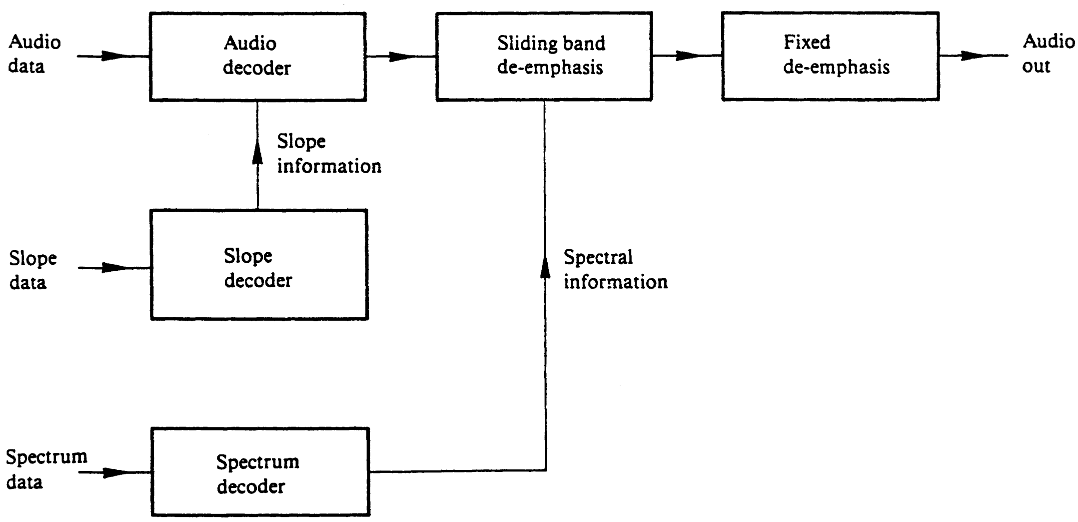


FIGURE 1 – *Adaptive delta modulation decoder*

3.2.2 *Sliding band de-emphasis (Fig. 2)*

The de-emphasis has the variable frequency response:

$$\left[\frac{10sT_1}{1+sT_1} + \frac{1+sT_2}{1+sT_3} \right]^{-1}$$

where $T_2 = 5$ µs, $T_3 = 50$ µs, and T_1 is variable under the control of the spectrum data.

Note that the first term in the bracket is a high-pass characteristic at frequency f_1 where:

$$f_1 = \frac{1}{2\pi T_1}$$

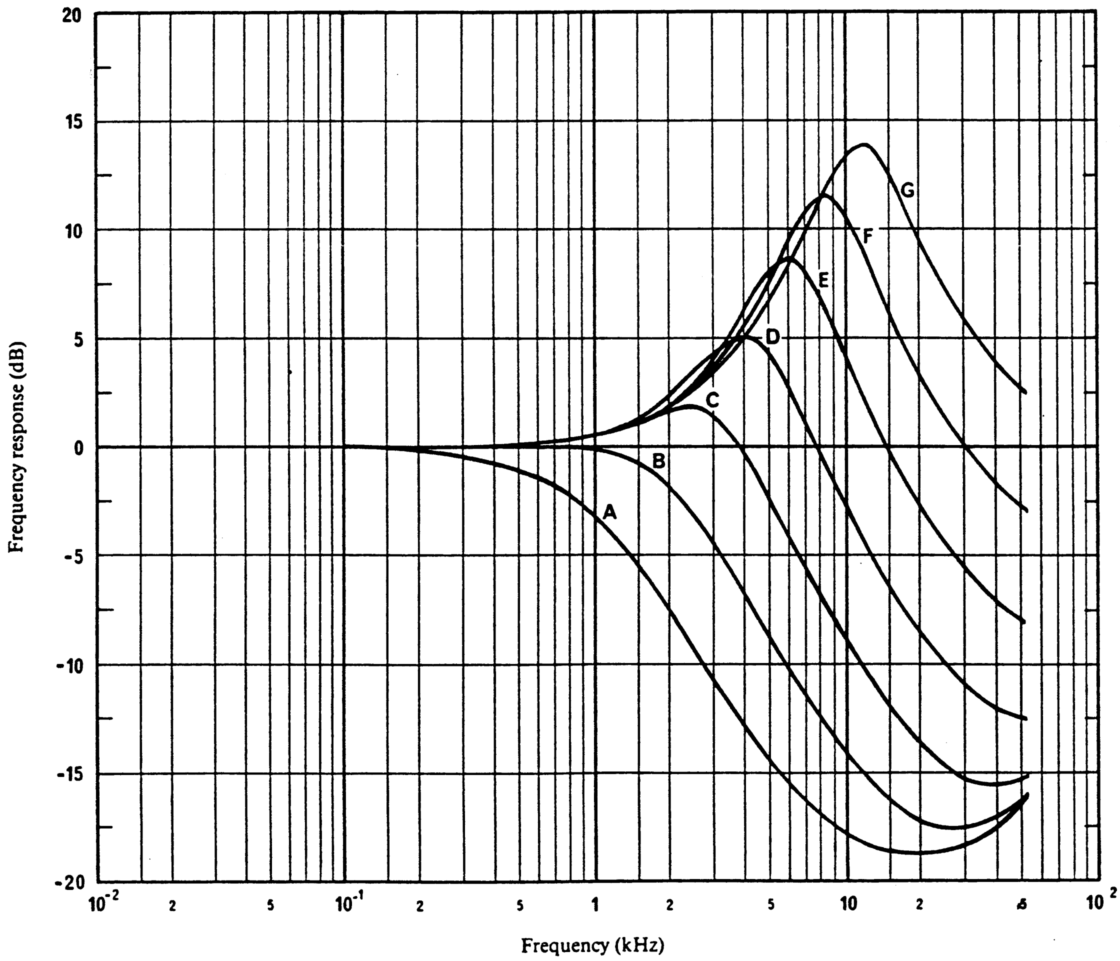


FIGURE 2 – *Adaptive delta modulation (ADM) de-emphasis filter*

3.2.3 *Spectrum decoder*

The de-emphasis control information is contained in a data bit stream at about 8 kbit/s, typically but not necessarily an integral sub-multiple of the audio data bit-rate. The mean level V_m of the bit-stream is derived by feeding the data via a three-pole low-pass filter with the characteristic:

$$(1 + sT_4)^{-3}$$

where $T_4 = 2$ ms.

If the data pulses have height V_p, a controlling parameter x is given by:

$$x = \frac{V_m}{V_p}$$

The output of the filter is fed into an exponentiator to produce a control signal to operate on the variable emphasis. The circuit constants are such as to provide the relationship:

$$f_1 = \frac{1}{2\pi T_1} = 4000\,(2^{10x}) \qquad \text{Hz}$$

With this definition, a change in x of 0.1 moves f_1 by one octave.

3.2.4 *Slope decoder*

The signal slope information is contained in another data bit stream at about 8 kbit/s. This bit stream is converted to a control signal by means of a low-pass filter and exponentiation exactly as in the spectrum decoder. The height of the pulses integrated in the audio decoder is linearly proportional to this control signal.

As in the spectrum decoder, the mean level V_m of the bit stream is derived using a three-pole low-pass filter with the following characteristic:

$$(1 + sT_4)^{-3}$$

where $T_4 = 2$ ms,

and is then fed into an exponentiator to provide the step or pulse size control signal V_{ss} for use in the audio decoder. This exponentiator has the characteristic:

$$V_{ss} = V_0(2^{10y})$$

where y is the normalized mean level of the pulses (as x above) and V_0 is a constant scaling factor to suit the audio decoder. With this definition, the pulse height changes 6 dB for each change in y of 0.1.

3.2.5 *Fixed de-emphasis*

The fixed de-emphasis is a single pole low-pass filter with the characteristic:

$$(1 + sT_5)^{-1}$$

where $T_5 = 25$ µs.

In consumer decoders no further output filtering is necessary.

REFERENCES

CCIR Documents

[1986-90]: a. 10/52 (Japan)

BIBLIOGRAPHY

FIELDER, L. D. [September, 1985] Pre- and post-emphasis techniques for audio recorders, *J. Audio Eng. Soc.*, Vol. 33, **9**.

FIELDER, L. D. [October, 1985] Audible modulation noise in floating point conversion systems, *J. Audio Eng. Soc.*, Vol. 33, **10**.

FIELDER, L. D. [October, 1985] Assessing the low level sound quality of digital audio systems. Presented at the 19th Convention of the Audio Engineering Society, Preprint No. 2268.

TODD, C. and GUNDRY, K. [March, 1984] A digital audio system for broadcast and prerecorded media. Audio Engineering Society, Preprint No. 2071, Paris, France.

TODD, C. and GUNDRY, K. [1984] A digital audio system coding method for DBS, cable and terrestrial broadcasting. Conf. Publ. 240, Ninth International Broadcasting Convention (IBC 84), Brighton, UK.

ANNEX I

EMPHASIS TECHNIQUES FOR HIGH-QUALITY PCM SOUND SIGNALS

In the technique of digital PCM sound coding, pre- and de-emphasis methods are to be used for several purposes:

– in the case of digital audio signal processing, the emphasis reduces the subjective perceptibility of the quantizing noise, especially for the so-called programme-modulated noise with companded coding;

– in the case of both companded and linearly coded audio signal processing, emphasis may reduce the impairment due to bit errors at low C/N values.

Two different characteristics have been proposed for both companded and linearly coded systems. The first pre-emphasis system is based on CCITT Recommendation J.17 (with an insertion loss of 6.5 dB at 0.8 kHz) and the second on the characteristic of 50/15 μs time constant. Characteristics of these are depicted in Fig. 1 of Recommendation 651.

Results of investigations into linear coding [CCIR, 1982-86a] and linear and companded coding [CCIR, 1982-86b and c] clearly show the advantages of using pre-emphasis in most cases. It has been observed that both systems have the quality of reducing the quantizing noise, with a slight advantage for 50/15 μs pre-emphasis in the case of non-companded operation. Concerning the risk of overloading, the documents [CCIR, 1982-86a, b and c] reflect different points of view. No clear difference between these two systems has been noted on impairments due to bit error at low C/N levels.

Results of theoretical assessments [CCIR, 1982-86b] and subjective tests [CCIR, 1982-86d and e] carried out for companded systems, based on the pre-emphasis characteristics shown in Fig. 1 of Recommendation 651, show that a significant reduction of the subjective audibility of the programme modulated noise could be achieved by applying CCITT Recommendation J.17 pre-emphasis. However, results of other subjective tests [CCIR, 1982-86f and g] showed no significant difference. In [CCIR, 1982-86f], insertion loss adjustment in each system was performed in order to avoid overloading.

As far as digital coding of sound programme signals is concerned, CCITT Recommendation J.17 (with an insertion loss of 6.5 dB at 0.8 kHz) is widely used for transmission purposes, as specified by CCITT Recommendation J.41, and the 50/15 μs pre-emphasis system is widely used for consumer applications.

REFERENCES

CCIR Documents

[1982-86]: **a.** 10/21 (EBU); **b.** 10-11S/201 (EBU); **c.** 10-11S/139 (Japan); **d.** 10/269 (Germany (Federal Republic of)); **e.** 10-11S/206 (France); **f.** 10-11S/205 (Japan); **g.** 10-11S/207 (Canada).

ANNEX II

SUBJECTIVE TEST RESULTS ON ADAPTIVE DELTA MODULATION AND OTHER CODING METHODS FOR HIGH-QUALITY SOUND

1. Subjective measurements conducted in Australia

The Australian Broadcasting Corporation (AuBC) has conducted subjective tests on the adaptive delta modulation (ADM) system [AuBC, 1985] in accordance with Recommendation 562 to relate subjective audio quality to bit error ratio, with the eventual objective of relating this to C/N ratio.

The supplied ADM equipment had a random error generator card to introduce errors into the data stream in a known switch-selectable ratio. The accuracy of the switch settings was checked and found to be accurate within normal statistical variations.

A carefully controlled and measured environment was used to replay the material in sequences according to Recommendation 562.

For evaluation of the audio part of the system, seven different segments of material were chosen from compact disc or original 16 bit PCM recordings and recorded in sequence in 16 bit PCM on a videotape machine for subsequent replay. The seven programme segments are listed in Table I.

TABLE I – *Reference programme segments*

Item (and curve) No.	Description
1	Male solo
2	Male spoken word
3	Piano
4	Flute, then boys' choir
5	Female spoken word
6	Orchestral
7	Mixed modern group

Figure 1 indicates that for a given BER, the impairments vary widely but consistently with the programme material, but that a mean impairment rating of 3.5 is achieved for a BER around 10^{-3}.

The overall impairment rating of 3.5 for a BER of 10^{-3} was thought to be equivalent to a C/N less than 9 dB. However, this would be degraded when using four levels of data as proposed for Australia. It was not possible to make measurements on this mode of operation, but results suggest a degradation of approximately 1 dB for the outer levels of data and approximately 1.8 dB for the inner levels. This may have implications as to which programmes should be carried on which data channels. However, facilities have not been available to the AuBC to corroborate these figures.

Independent correlation is yet to be established between C/N and BER obtainable from receivers in the field.

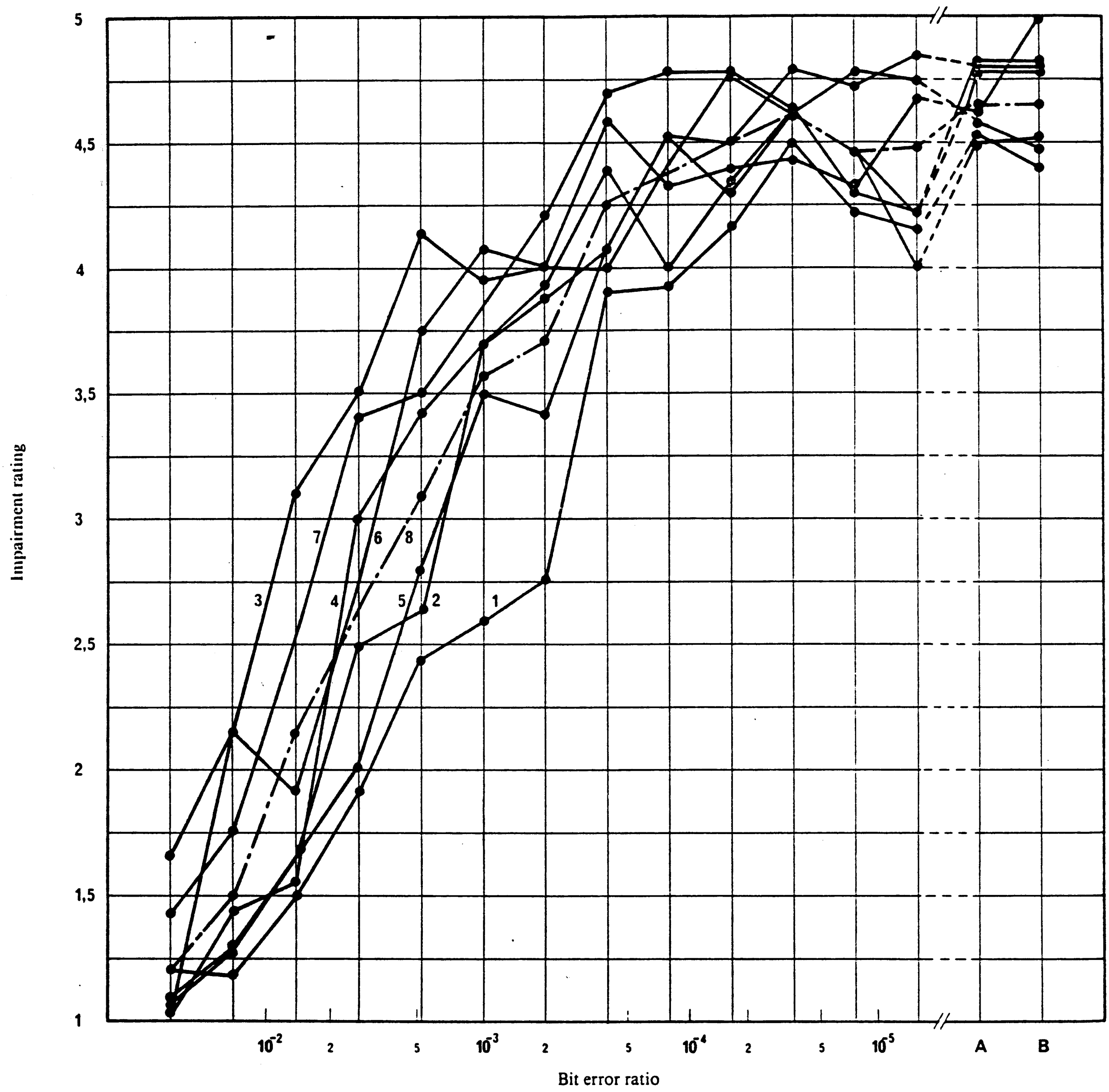


FIGURE 1 – *Impairment rating versus switch position, item by item*

Note 1. – Curves No. 1-7 are from programme selections indicated in Table I. Curve 8 is for overall (average) result.

Note 2. – In scale position A, signal is transmitted through Dolby ADM system without errors.

Note 3. – In scale position B, signal comes direct from analogue source (e.g.: PCM on VCR).

2. Subjective measurements conducted in Canada

2.1 *Subjective asssesment of a selection of ADM, linear and near-instantaneous codes*

In the process of identifying the relative merits of the various proposed coding schemes, a test programme was conducted in Canada.

The following coding schemes were covered by the measurement programme:

- 10-14 semi-instantaneous companding (NICAM 3) with CCITT Recommendation J.17 pre-emphasis;
- 10-14 instantaneous companding (A-law) with CCITT Recommendation J.17 pre-emphasis;
- 14 bit linear coding with CCITT Recommendation J.17 pre-emphasis;
- adaptative delta modulation (sampling rate = 330 kHz); and
- adaptative delta modulation (sampling rate = 204 kHz).

The first three coding schemes were modelled on a computer working in conjunction with a vectorial processor.

The pre-emphasis applied to the signal was changed through digital filtering. The sampling frequency was also changed to 32 kHz through digital processing. Companding and truncation to generate the proper number of bits was then performed to finally obtain the companded digital stream. Proper re-formatting was made to store the test sequence on video tape.

This process was not used for the ADM system which could not be modelled on the computer. Conversion to analogue signal was necessary before processing by the ADM. At the output of the ADM the signal was converted back from analogue to 16 bits/sample for insertion in the recorded test sequence.

The source material was either read from compact laser discs or generated by the computer itself:

- Cardozo, Pájaro Campana (Indian Harpsichord)
- Orff, Carmina Burana (Lyrics)
- synthetic gong generated at 44 kHz simulating the sound of a triangle on the first 4 notes of "Frère Jacques" (4.2 kHz, 4.7 kHz, 5.3 kHz and 4.2 kHz). (Attack: T = 3 ms, decay: T = 300 ms.)
- synthetic gong generated at 44 kHz simulating a bass guitar on the first 4 notes of "Frère Jacques" (65.4 Hz, 73.4 Hz, 82.4 Hz and 65.4 Hz). (Attack: T = 3 ms, decay: T = 300 ms.)

Tests were made according to the suggested CCIR comparison method (Recommendation 562) where sequence A-B is presented twice and then 15 s is left for scoring also according to the CCIR comparison scale. Audiometric measurements were taken with each subject before the subjective tests. On average, three listeners were accommodated in a well-calibrated listening room. A total of 25 listeners performed the tests.

All comparisons were repeated in reverse order to have a measure of the listener's consistency. Accordingly, 4 listeners were discarded from the analysis of the results. Furthermore, when the two scores of a same listener for the same comparison differed by more than two grades, these were discarded.

Due to uncontrollable vibrations and rattles induced in the listening room by the high-level low frequency synthesized gong, the test sequence was also assessed by 8 listeners using headphones. The results of this group are used for the low frequency gong.

In addition to the mean opinion score and the standard deviation obtained from the filtered data, a study of the statistical significance of the results was conducted using the one tail Student-t distribution. The level of confidence that, on average, the normal population will prefer codec "A" to codec "B"* was obtained.

Table II summarizes the results. The mean opinion score expressed on the −3, 0, +3 scale, the standard deviation as well as the level of confidence that "codec A is better than codec B" are given in each case. In the table, the 10-14 near-instantaneous companding as described in Report 953 is noted as NICAM 3.

---

* "Codec" = "coder-decoder".

TABLE II – *Results of the subjective tests conducted in Canada*

Codec "A"	Codec "B"	Harpsichord $\bar{X}$ / S / % Conf	Lyrics $\bar{X}$ / S / % Conf	HF gong $\bar{X}$ / S / % Conf	LF gong $\bar{X}$ / S / % Conf
NICAM 3, CCITT Rec. J.17	ADM (330 kHz)	0.1 0.7 88%	0.2 0.9 91%	−0.1 1.3 28%	1.1 0.8 99.95%
NICAM 3, CCITT Rec. J.17	ADM (204 kHz)	0.0 0.8 50%	0.2 0.9 92%	1.8 1.0 99.999%	1.4 0.5 99.995%
ADM (330 kHz)	A-law, CCITT Rec. J.17	0.0 0.6 50%	−0.2 0.8 6%	0.9 1.4 99.99%	−1.2 0.5 0.01%
NICAM 3, CCITT Rec. J.17	A-law, CCITT Rec. J.17	0.0 0.7 50%	−0.1 1.0 27%	0.8 1.1 99.975%	0.1 0.3 88%
NICAM 3, CCITT Rec. J.17	14 bits linear, CCITT Rec. J.17	−0.1 0.7 18%	−0.1 0.8 19%	−22 0.8 0.001%	0.1 1.0 60%

$\bar{X}$: mean opinion score on a scale −3 to +3 as per Recommendation 562.
S: standard deviation.
% Conf: statistical confidence level that codec "A" is preferred to codec "B".

As can be seen from the table, no significant difference was found between these two encoding systems when the ADM was sampled at 330 kHz except in the case of the critical low frequency gong where a noise spectrum shaping was perceived. This was mostly noticeable during the decay time of the gong. When the ADM was sampled at 204 kHz, a degradation in the reproduction of the high frequency gong could also be perceived, giving worse performance than the NICAM 3. For these two cases, however, no significant difference could be perceived between the two codecs for normal programme material.

2.2 *Quality assessment of a near-instantaneous differential code and a near-instantaneous PCM code*

An objective and subjective evaluation of two coding schemes, proposed in Report 1075 ——— for HDTV broadcasting by satellite, was done in Canada. The coding schemes considered were:

- 15-to-8 Near-Instantaneous Companded DPCM (NI-DPCM) proposed for MUSE;*
- 14-to-10 Near-Instantaneous Companding (NICAM) as per Recommendation 651.

* The NI-DPCM scheme used in MUSE is based on a leakage factor of 0.9375 as compared with a leakage factor of 0.975 used in the simulations reported here. Simulations with leakage factor values between 0.95 to 0.99 showed only minor differences in signal-to-noise ratios.

2.2.1 Objective evaluation

Both coding schemes were simulated on a computer in accordance with the descriptions given in Report 953-1 for the NICAM and in Report 1075 ———— for the NI-DPCM. A leak factor of 0.975 was used in the simulated NI-DPCM. The following audio test signals were applied to each encoder:

a) variable frequency sinusoidal signal
b) low and high frequency synthetic sounds
c) natural harp

The input test signals to the encoders consisted in 32 kHz digital audio samples uniformally quantized to 15 bits/sample for the NI-DPCM encoder and to 14 bits/sample for the NICAM encoder. For each test signal and for each encoder, signal-to-noise ratios were calculated over successive blocks of 256 samples (8 msec) and plotted as a function of time.

The variable frequency sinusoidal signal consisted in a sequence of ten sinusoids of constant amplitude and variable frequency ranging from 30 to 15360 Hz. Figure 2a shows an excerpt of this test signal and the SNR curves obtained are plotted in Fig. 2b and 2c. It can be seen from Fig. 2c that an instantaneous quantizing scheme such as the NICAM has an SNR which is independent of the frequency. A differential coding scheme like the NI-DPCM will yield better SNR values at low frequencies (up to 1 kHz). However, the SNR values deteriorate at higher frequencies, to become smaller than those obtained with the NICAM.

The synthetic sound test signal consisted in the first four notes of "Frère Jacques" (Do-Re-Mi-Do) synthesized both at low (60-80 Hz) and high frequency (4-6 kHz). Figure 3a shows the resulting signal shape in the time domain and the SNR curves obtained for the low frequency sound are plotted in Fig. 3b. The superiority of the differential coder (NI-DPCM) is apparent for a high level low frequency signal despite its lower number of bits-per-sample (8 against 10 for NICAM). Both coders exhibit decreasing SNR for decreasing signal levels. The SNR curves obtained with the high frequency synthetic sound are shown in Fig. 3c. The SNR values obtained for the NICAM are superior, for high level signals, to those obtained with the NI-DPCM.

The natural harp sound, shown in Fig. 4a, exhibits a very large dynamic range as well as large variations of level over short time periods. The SNR curves obtained with this test signal are shown in Fig. 4b and 4c. These two curves are comparable with, perhaps, a slight advantage for the NI-DPCM over the NICAM when the input signal reaches the peaks of its dynamic range.

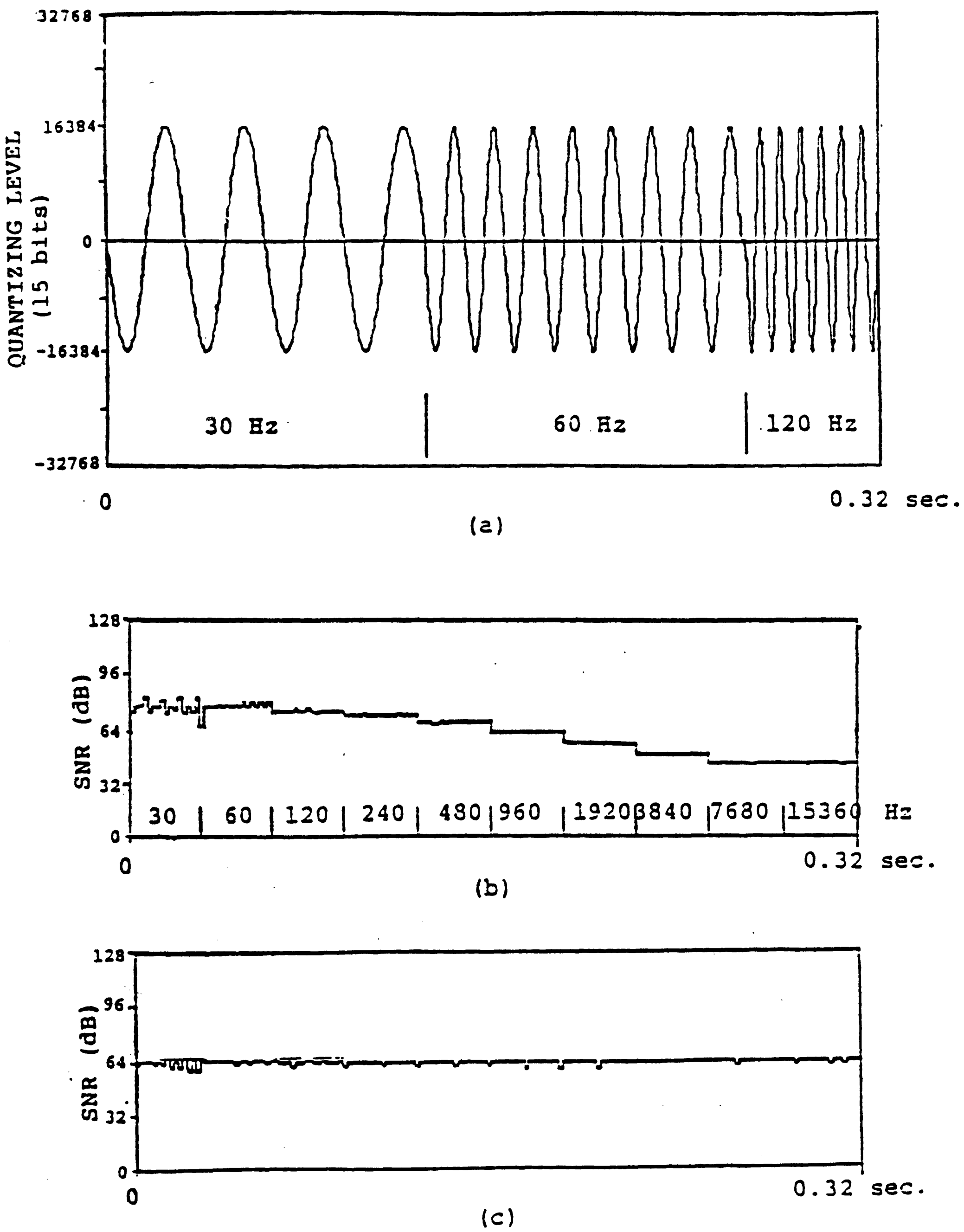
2.2.2 Subjective evaluation

An informal listening test using headphones was conducted with 4 sound broadcast engineers and 2 student engineers as listeners. Tests were made according to the suggested CCIR comparison method (Recommendation 562-2) where sequence A-B is presented twice and then about 15 sec. is left to the listener for scoring between -3 (coder A much worse than coder B) to +3 (coder A much better than coder B).

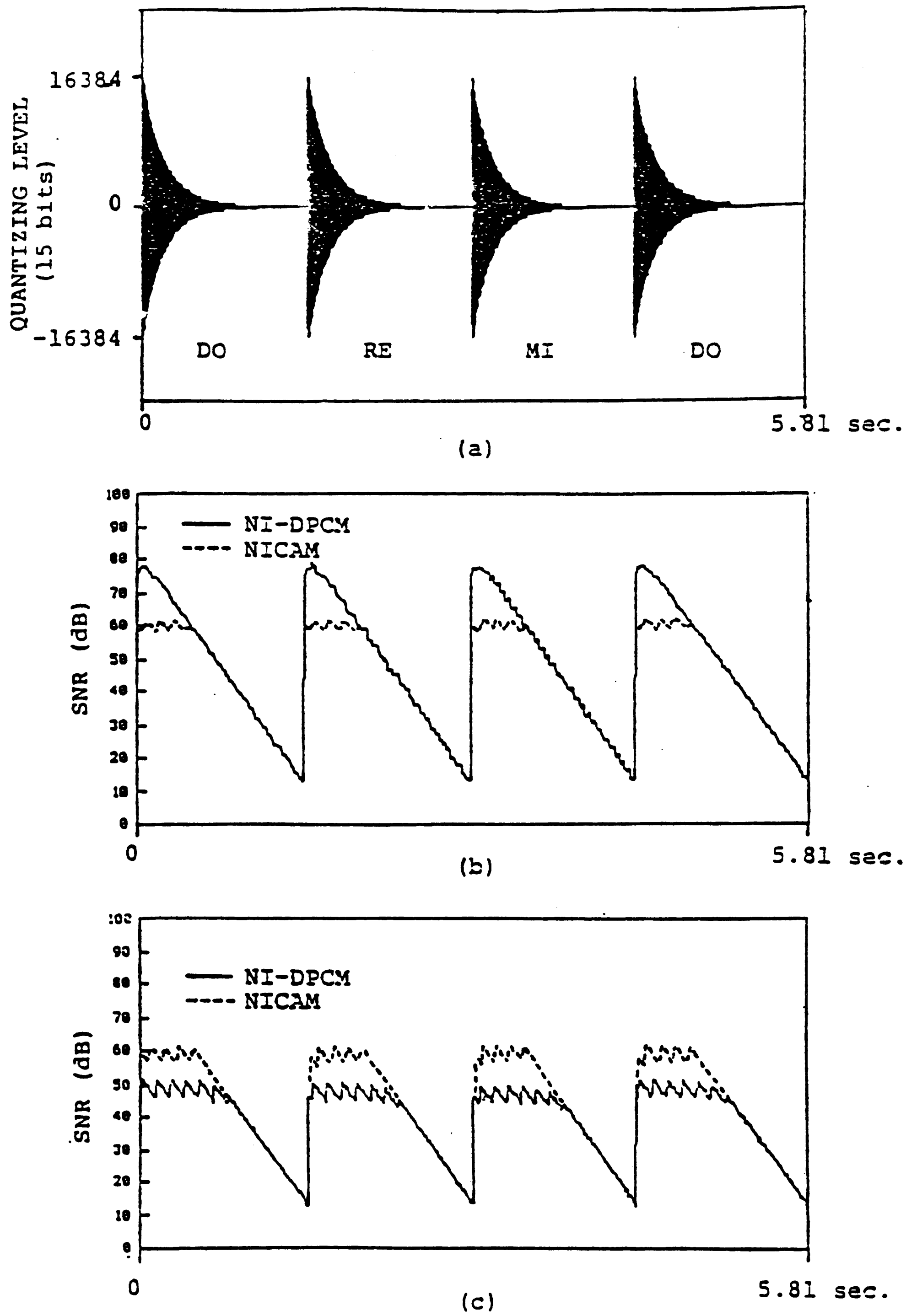
The test material consisted in excerpts from a harp solo, an organ solo, a female solo and performances from a chamber music ensemble, a jazz quartet and a pop music band. The low and high frequency synthetic sounds described in section 2.2.1 were also included in the test material. The 50/15 us pre-emphasis/de-emphasis law described in Recommendation 651 was used with both NI-DPCM and NICAM encoders.

The listening test, although informal, yielded consistent results. The listeners found no significant differences between the NI-DPCM and the NICAM encoders for the six natural music excerpts. The average opinion score for these six sequences was 0.2 in favor of NI-DPCM with a standard deviation of 0.8.

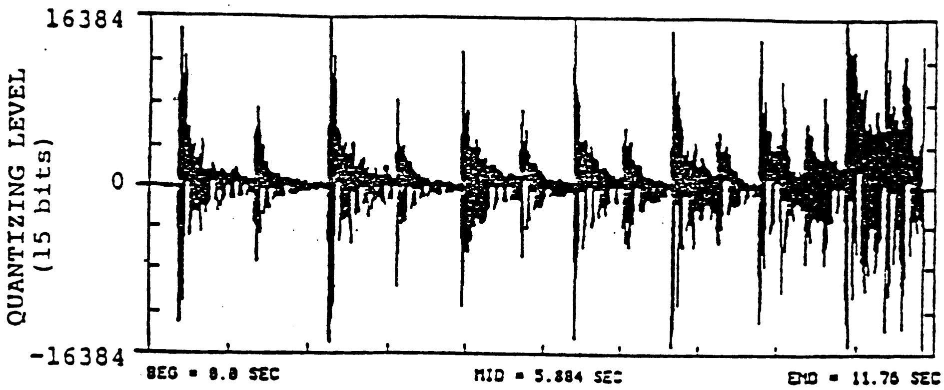
The listening panel clearly preferred the NI-DPCM encoder for the low frequency synthetic sound (average score of 2.2, standard deviation of 0.7) and the NICAM for the high frequency synthetic sound (average score of -1.8 with a standard deviation of 0.7). These results are in agreement with the objective evaluation results of section 2.2.1 where the NI-DPCM encoder had yielded better SNR values than the NICAM for low frequency signals but poorer SNR at high frequency."



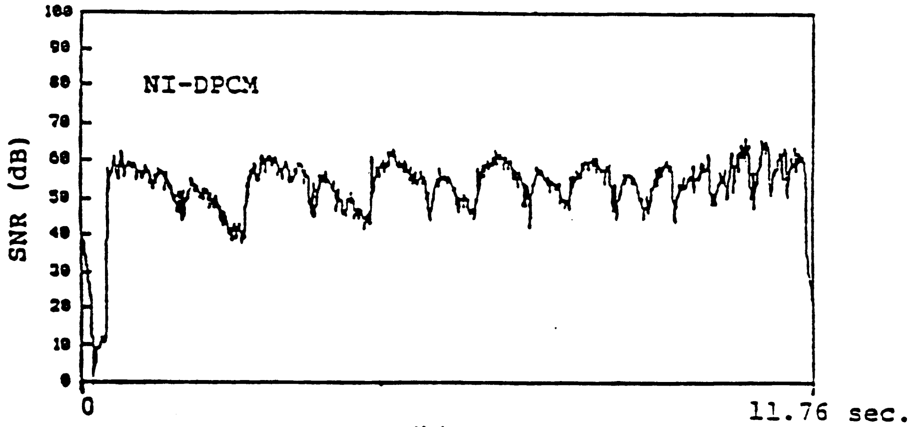
Figure[2] Variable frequency sinusoidal signal
a) test signal excerpt
b) NI-DPCM: SNR vs frequency of sine wave
c) NICAM : SNR vs frequency of sine wave



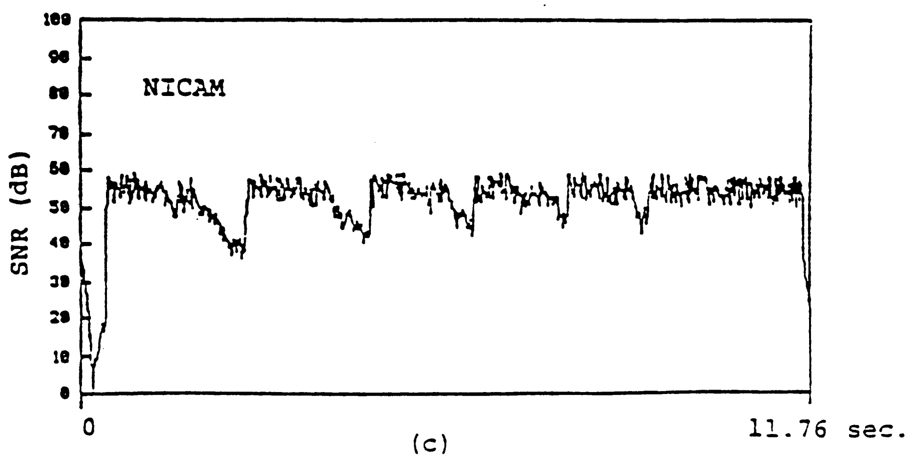
Figure[3] Low and high frequency synthetic sounds
a) test signal waveform
b) SNR curves for low frequency sound
c) SNR curves for high frequency sound



(a)



(b)



(c)

Figure[4] Natural harp sound
a) test signal waveform
b) SNR curve for the NI-DPCM coder
c) SNR curve for the NICAM coder

3. Subjective measurements conducted in the United States

Carefully controlled listening tests using trained auditors were conducted to evaluate the adaptive delta modulation (ADM) system described in Report 953-1. The experiment tested and compared the ADM system (sampling rate = 300 kHz) against a full dynamic range compact disc.

Twenty-four well-motivated and highly-qualified listeners, whose hearing was clinically normal as confirmed by an audiometric evaluation performed by a licensed and certified audiologist, were hired for these tests. The listeners' ages ranged from 16 to 27 years to ensure the probability of normal hearing. They were all experienced or expert listeners who were instructed to "vote" for the system which sounded superior in quality. No further details of what to listen for were provided. Fifteen listeners were male, nine were female, and they were paid in order to enhance motivation.

The listening conditions (environment) were kept in close agreement with those in Recommendation 562. The tests were conducted in a listening studio which had carpet and drapes, had a mid-frequency reverberation time of approximately 0.25 second, and was 16 x 22 feet. The loudspeakers were placed approximately 10 feet apart. A single listener at a time was tested.

The listening level in the center of the room was set to 85 dB on loud music passages. This level was chosen because at lower levels, differences between systems might not be perceptible.

The 10 CD selections, which are shown in TABLE III, were chosen from the European Broadcasting Union (EBU) Sound Quality Assessment Material (SQAM) Compact Disc (CD) as critical test material:

TABLE III - Programme Segments

Electronic Tune -- Frère Jacques
ABBA -- Pop Music
Triangles -- Single Instrument
Grand Piano -- Single Keys & Brief Musical Selection
Bells -- Single Instrument
Eddie Rabbitt -- Country Music
Soprano -- Solo
Wind Ensemble -- Brief Musical Selection
Xylophone -- Single Keys & Musical Selection
Male Speech -- German

These stereo selections used for test purposes have wide variations in dynamic range, spectrum, and temporal characteristics. They represent typical and atypical sound programme material. They were chosen to be especially problematical for the ADM type of processing under test, and to reveal coding artifacts. System performance flaws which might be audible include: A/D and D/A linearity, aliasing distortion, bit-rate reduction and programme-modulated noise.

The excerpts were 20 to 140 seconds in duration, and the total subjective-test time was 1/2 hour.

The classic two-alternative forced-choice (2 AFC) psychophysical method was employed. (The method of Recommendation 562 was not appropriate since this was a signal detection task, not a quality scaling task.) The ten selections were presented four times each, in two random orders. Listeners switched back and forth between systems in a pair and then voted for the one of "superior sound quality" immediately after each selection ended. Approximately five seconds separated the segments.

The initial-test results are presented in TABLE[IV] such that for each selection both the number of judgements and the percentage which that number represents are reported for each audio selection. This format allows an examination of the influence, if any, of the audio segment selection on the choice of the listeners and it also allows a direct comparison of the two audio paths. When there is no perceptible difference, this test produces results which are randomly distributed with a mean of N/2 (50%), and a standard deviation of $\sqrt{N/4}$ (where N is the number of trials).

Each listener made 40 judgments, four votes on each of ten selections. The total group of twenty-four, therefore, made judgments of each selection for a total of 960 judgments. Random selection would yield (per selection) 48 votes (50%) for the CD player, with a standard deviation of 4.9 votes (5.1%)

The overall initial-test result (Table [IV]) is 52.5% for ADM and 47.5% preference for the CD player, a very slight difference, if any. From this overall view alone, it could be concluded that the compander is acoustically transparent. However, there were five selections in which a preference more than 5% different from 50% occurs (45 to 55%): the electronic tune, 55.2% for CD; ABBA, 65.6% for ADM; the triangles, 61.5% for CD; the soprano solo, 59.4% for ADM; and the xylophone, 62.5% for ADM, and there appeared to be 3 to 5 listeners who were the primary cause of this result.

Since a few listeners were able to detect differences on these critical selections (although there was no preference for either audio path) a follow-on study was conducted to determine if small audio level differences might be responsible for the results. Level matching of the two paths (direct CD and CD via ADM codec) was improved from approximately 1.0 dB to better than 0.5 dB). All test conditions were kept the same. The results are shown in TABLE V.

The ADM codec, after the additional level matching adjustments, became even more difficult to distingush from the direct CD signal path by the expert listners. Three of the five audio program segments shifted approximately 10 percentage points toward true random results (50/50) and ended at 44, 50, and 52%. The results for the remaining two selections did not change appreciably (2 to 3 percentage points).

A carefully controlled test has shown that experienced and expert listeners showed no preference for the sound of critical programme material reproduced from a direct compact disc source over that of the CD played through the ADM codec under the test conditions described above.

TABLE IV

PILOT LISTENING TEST RESULTS
AUGUST 1988

VOTES FOR CD PLAYER
24 test subjects
96 forced-choice judgments per selection

	TOTAL VOTES	PERCENT
1 - Electronic Tune #7	53	55.2
2 - ABBA #69	33	34.4
3 - Triangles #32	59	61.5
4 - Grand Piano #39	47	49.0
5 - Bells #34	48	50.0
6 - Eddie Rabbitt #70	49	51.0
7 - Soprano #44	39	40.6
8 - Wind Ensemble #66	47	49.0
9 - Xylophone #36	36	37.5
10 - Male Speech-German #54	45	47.0
TOTALS	456	47.5

The standard deviation of 96 random events = 4.9 (5.1%).

TABLE V

FOLLOW-ON LISTENING TESTS RESULTS
SEPTEMBER 1988

VOTES FOR CD PLAYER
20 test subjects
160 forced-choice judgments per selection

	TOTAL VOTES	PERCENT
1 - Electronic Tune #7	86	53.8
2 - ABBA #69	70	43.8
3 - Triangles #32	80	50.0
4 - Soprano #44	83	52.0
5 - Xylophone #36	65	40.6
TOTALS	384	48.0

The standard deviation of 160 random events = 6.3 (4%).

REFERENCE

AuBC [May, 1985] Australian Broadcasting Corporation Engineering Research and Development Report No. 139. Subjective Tests of Adaptive Delta Modulation with B-MAC for HACBSS.