Proceedings of the Institute of Acoustics

DATA COMPRESSION ALGORITHMS FOR HIGH DEFINITION DIGITAL AUDIO USING MULTI-PULSE ADAPTIVE SUB-BAND CODING (MASC)

R.K.C. Tan (1) & Dr M.O.J. Hawksford (2)

(1) & (2) University of Essex
Department of Electronic Systems Engineering
Audio Research Group
Wivenhoe Park, Colchester CO4 3SQ, UK

Abstract

A system of data compression is proposed for use with an oversampled and noise shaped High Definition Digital Audio (HDDA) format (20 bits/88.2 kHz) which enables the performance parameters of a linear PCM format (16 bits/44.1 kHz) to be extended. An efficient coding algorithm based upon the psychoacoustic characteristics of the human auditory system is incorporated. Research has considered four fundamental coding techniques, to determine their suitability with sub-band coding, ie. APCM, ADPCM, ADM and LPC. Proposed as the Multi-pulse Adaptive Sub-band Coding (MASC) technique, the HDDA system combines the application of linear predictive coding (LPC) and sub-band coding (SBC).

0 INTRODUCTION

An extension of the existing digital audio format (16-bit/44.1kHz) to a High Definition Digital Audio (HDDA) format is necessary in order to reproduce sound closer to that of the original musical performance. This would require a 20-bit system to increase the dynamic range to 120dB, in which the human ear can operate. To obtain a better transient response from the effect of low-pass filtering at 20kHz, the available bandwidth has to be increased. A higher sampling rate of 88.2kHz is suggested.

Our proposed oversampled and noise shaped HDDA format results in a bit rate increment from 705.6kbits/s per monaural channel to 1.764Mbits/s per monaural channel. Consequently efficient data compression scheme must be applied to substantially reduce the high data rate associated with such systems. This has strong economic and design consequences in the storage and transmission areas; it allows the preservation of HDDA master recording sound quality onto the existing 5-inch CD medium without compromising the playing time of approximately 80-minutes. Alternatively, CD quality music signals can be transmitted through ISDN channels at 128kbits/s per channel. In Digital Audio Broadcasting (DAB), more high-quality stereo programmes can be multiplexed into a finite electro-magnetic spectrum. Compatibilities between the HDDA systems and the existing systems can also be retained.

An important consideration when performing data compression upon audio signals is a reduction in bit rate without compromising PCM sound quality. Conventional techniques have been successful in achieving these objectives using system theories which remove redundant parts of the audio signal [1]. However, the degree of compression required of the HDDA data cannot be achieved by using such system theories alone. Psychoacoustic characteristics of the human ears must also be considered, coding bits of information which match the human hearing characteristic [2-6].

Our proposed Multi-pulse Adaptive Sub-band Coding (MASC) data compression technique considers auditory masking and combines the application of multi-pulse excitation Linear Predictive Coding (LPC) with Sub-band Coding (SBC). The audio spectrum is split into four bands by considering the way sound are perceived.

This paper attempts to give a concise discussion of the data compression methods investigated. Section 1 highlights research which has considered four fundamental coding techniques and their suitability for use with SBC, where both waveform coding and parametric coding techniques were considered [7]. Since the characteristics of the sub-band signals are considerably different from those of broadband audio signals, techniques developed for encoding of broadband signals do not necessarily lead to good results with sub-band signals. Section 2 describes the proposed MASC design and presents the experimental results obtained. Finally, Section 3 summarises the findings and provides a critical view of the proposed MASC system.

1 SELECTION OF SUB-BAND CODING TECHNIQUE

1.1 Adaptive Pulse Code Modulation (APCM)

From the results of the APCM coder error spectrum analysis shown in Figure 1(A), we noticed that the re-quantization error is actually white noise. At the higher frequency regions where the spectral energy is lower, there is basically no noise masking effect as the noise energy is higher than the actual signals.

Figure 2(A) shows the results obtained from objective test measurements. We notice that the performance of the APCM coder decreases only very slightly (maximum of about 4dB difference) for audio signals which have higher amplitude levels (drum and pop music signals). This is due to the coarser quantization levels used at higher amplitude levels. On the other hand, signals which have lower amplitude levels (piano and male vocal) perform marginally better with the APCM coder due to the finer quantization levels used.

Signals with low sample-to-sample correlation are best performed by APCM coding, a point noted for sub-band signals at the higher frequency regions. At these frequencies, signals are uncorrelated from sample-to-sample and APCM does not depend on differential coding between adjacent samples. Hence, the reason for the better performance. However, it gives a poorer performance at the lower frequency regions.

1.2 Adaptive Differential Pulse Code Modulation (ADPCM)

Quantization error can effectively be reduced by differential coding since the difference between adjacent samples has a smaller variance than for PCM samples as shown in Figure 1(B) for ADPCM coder. Again we notice that the re-quantization error is white noise from the error spectrum analysis. Hence no noise masking effect was observed in the higher frequency regions.

Figure 2(B) shows the results obtained from objective test measurements using the six audio signals utilized in the APCM tests. As expected, we can see that the performance of the ADPCM coder increases for audio signals which exhibit a relatively high sample-to-sample correlation; male speech, piano and guitar. For audio signals with low sample-to-sample correlation (violin, pop music and drum) ADPCM performance is poorer.

ADPCM is a more suitable technique to use with sub-band signals than APCM in the lower frequency bands of the music spectrum since a relatively high sample-to-sample correlation exists. However, at the higher frequency regions the performance deteriorates.

1.3 Adaptive Delta Modulation (ADM)

Signals which exhibit very high sample-to-sample correlation can be coded to one-bit with ADM. Despite its simplicity, re-quantization noise and granular noise were subjectively very obvious during listening tests. Increased sample-to-sample correlation was achieved by oversampling the input signals [8]. Although a considerable improvement in the subjective signal quality was obtained, re-quantization noise was still audible above 3kHz at 8-times oversampling rate as shown in Figure 1(C). A very high oversampling rate (up to 256-times) is required to obtain adequate subjective quality [9-10].

Figure 2(C) shows that signals with high sample-to-sample correlation perform better with ADM (piano, guitar and male speech). Likewise, signals which are not well-correlated sample-to-sample (drum, violin and pop music) perform rather poorly. However, by utilizing the effect of oversampling we can see that the performance of all six audio signals are almost equal (less than 2dB difference). We can conclude that data compression with an ADM coder requires a high oversampling rate in order to achieve high sound quality.

An interesting factor when performing multiple narrow sub-band filtering is that the signal in each band would be highly oversampled if decimation was discarded. However, as soon as the number of sub-bands exceeds the number of bits (eg. 16-bits) in the linear PCM system, the final data rate will be higher. This contradicts the main objective of data compression.

1.4 Linear Predictive Coding (LPC)

The basic idea of LPC [11] is exactly the same as one form of ADPCM which was introduced earlier. Instead of just one past value of the signal used to form the prediction, LPC use several past values. The prediction error could be written as

$$\begin{aligned} d(n) &= s(n) - a_1 s(n-1) - a_2 s(n-2) - \dots - a_p s(n-p) \\ &= s(n) - \sum_{k=1}^{p} a_k s(n-k) \end{aligned}$$

where the multipliers a_k are adapted to minimize the error signal.
The equivalent z-transform relationship gives

$$\begin{aligned} D(z) &= S(z) - \sum_{k=1}^{p} a_k z^{-k} S(z) \\ &= \left(1 - \sum_{k=1}^{p} a_k z^{-k}\right) S(z) \end{aligned}$$

Rewriting the signal in terms of the error,

$$S(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} D(z)$$

From the above equation, the signal during synthesis is represented by an all-pole filter with the predictive error $D(z)$ identified as an impulsive excitation as shown in Figure 3. This raises a possibility of parametrizing the error signal by its frequency and amplitude instead of coding it sample-by-sample. These two parameters are relatively slow-varying and hence signals can be transmitted at a rate much lower than the sampling frequency. An extremely low data rate can therefore be achieved.

The excitation for LPC synthesis is usually classified into two separate groups, voiced and unvoiced. For voiced signals, the excitation is a quasi-periodic pulse train with delta functions located at pitch period intervals. For unvoiced signals, the excitation is white noise. An LPC-based coder is considered to be unsuitable for use with broadband audio signals due to the complexity of the audio waveforms. Accurate separation of audio signals into two classes, voiced and unvoiced, (already difficult to achieve in speech) is not possible.

Newer LPC-based coders with multi-pulse excitation, however, can compensate for quantization errors in the LPC parameters by appropriately adjusting the computed excitation. An analysis-by-synthesis procedure described by Singhal and Atal [12] to compute the excitation was simulated as shown in Figure 4. In multi-pulse excitation, the signal is not classified into voiced and unvoiced frames. Instead the excitation consists of a number of impulses for each frame, so that the final result is as close to the original as possible.

The excitation generator produces a sequence of pulses as the input to the linear filter. The resulting error signal between the output of the linear filter and the original signal is frequency weighted and fed back to the excitation generator. The excitation pulses are computed by the excitation generator on a frame-by-frame basis by minimizing the energy in the weighted error signal. The desired output quality is determined by the number of pulses placed in each frame of input. With more pulses per frame, the output quality increases as shown in Figure 5.

Figures 1(D) shows the error spectrum of the multi-pulse LPC coded signal. We can see that the re-quantization noise is non-white compared to APCM and ADPCM coding due to the weighting filter. This has the effect of concentrating the error energy under the portions of the frequency spectrum where the audio signal has higher energy, hence achieving the desired masking phenomenon.

Figure 2(D) shows the performance of various audio signals, as the number of pulses per frame is increased. In this experiment, a frame size of 10ms (441 samples) was used for the LPC analysis with the 24th order LPC filters. We notice that narrower bandwidth signals like piano, guitar and male speech perform well with the multi-pulse LPC coder. On the contrary, wider bandwidth signals like violin, pop music and drum give an inferior performance with multi-pulse LPC coder (max. 23dB difference). This is because both high and low pitched instruments are present in wider bandwidth audio signals. The LPC filter merely models the overall envelope of the audio waveforms, and hence, more excitation pulses are required in order to model higher bandwidth signals. With narrower bandwidth audio signals, the time-domain waveforms are near-sinusoidal and the LPC filter can model the overall envelope of the waveforms with fewer excitation pulses. Therefore, sub-band signals should perform better with multi-pulse LPC coder than APCM, ADPCM or ADM.

2 MULTI-PULSE ADAPTIVE SUB-BAND CODING (MASC)

2.1 4 x Sub-band Filters Design

In the MASC design shown in Figure 6, the music signal becomes more deterministic in each sub-band, behaving like a sine-wave which is very suitable for multi-pulse LPC coding. The system offers advantages over a broadband system due to the individual adaption of the masking threshold in each band according to its signal variance. Modulated re-quantization noise is also constrained to each band and cannot interfere with signals in any other bands. The advantage of this is that noise masking by the dominant signal is much more effective due to the reduction in the noise bandwidth.

The first band was obtained by low-pass filtering to preserve the dc content while the critical bands and upper band were obtained by band-pass filtering and high pass filtering respectively, as shown in Figure 7. The four bands were split from an audio frequency spectrum between 0Hz to 22.05kHz at a sampling rate of 44.1kHz. Figure 8 shows the frequency response of the four bands filters implemented.

2.2 Experimental Results

Firstly, the quality of the output signals from broadband multi-pulse LPC is compared with sub-band multi-pulse LPC. Setting the required overall bit-rate at 128kbits/sec per channel for a 44.1kHz sampling rate audio system is equivalent to a requirement of 28-pulses for broad-band coding [see APPENDIX, Section 5]. This is assuming that 16-bits are used to encode the pulse locations, amplitudes and the LPC filter coefficients. A frame size of 10ms (441 samples) per channel and 24th order LPC filters were used.

The performance of multi-pulse LPC suffers at higher frequencies from broadband coding as seen in Figure 9(A). If the audio spectrum is divided into sub-bands, signals from each band will have a narrower bandwidth, and hence the signals can be modelled more accurately without needing a large number of pulses. Figures 9(B)-(D) shows the effect upon 4-bands LPC coding as the number of pulses in each band is varied. It can be seen that as the number of pulses is increased in each band, the error is reduced within that band (re-quantization noise produced in one band cannot interfere with other bands). We can therefore allocate more pulses in the critical bands and in the upper band where noise masking is minimal. This coding technique is definitely more efficient than coding broad-band signals. Hence, MASC coding can offer data compression at a much lower bit-rate than broadband coding for constant sound quality.

In our research, we have managed to compress audio data to a bit-rate as low as 94.4kbits/s per channel with excellent results as shown in Figure 9 (D). This is equivalent to encoding audio signals with approximately 2-bits in linear PCM coding at 44.1kHz sampling rate. The bit-rate could be further reduced to less than half of the present value since less than 8-bits is sufficient to encode the multi-pulse LPC parameters.

3 CONCLUSION

Multi-pulse LPC coding can produce non-white re-quantization noise which adapts according to the signal energy content by the use of weighting filters. This has the effect of concentrating the error energy under the portions of the frequency spectrum where the audio signal has higher energy and hence noises will be more effectively masked. Multi-pulse LPC coders also perform better with narrow bandwidth signals.

The proposed Multiple Adaptive Sub-band Coding (MASC) algorithm has been described which considers the psychoacoustic characteristics of human hearing. Using the MASC coding techniques, we found that high quality sound can be produced at very low bit-rate by concentrating on coding the critical bands accurately and also by paying more attention to the upper frequency bands where noise masking is minimum. In our experiment, we have managed to compress the audio data to a rate as low as 94.4kbits/s per channel with excellent results using 16-bits to code the multi-pulse LPC parameters. A bit-rate halving can be further achieved since less than 8-bits is sufficient to encode the multi-pulse LPC parameters.

4 REFERENCES

[1] **C R Caine, A R English & J W H O'Clarey**, "NICAM 3 : Near-Instantaneously Companded Digital Transmission system for High-Quality Sound Programmes"
The Radio and Elect. Engr., Vol.50, No.10, (Oct 1980) pp 519-530.

[2] **B C J Moore**, *An Introduction to the Psychology of Hearing*
publisher Academic Press, (1989 edt).

[3] **L D Fielder**, "Evaluation of the Audible Distortion and Noise Produced by Digital Audio Converters"
J. Audio Eng. Soc., Vol.35, No.7/8, (Jul/Aug 1987) pp 517-535.

[4] **E Zwicker & U T Zwicker**, "Audio Engineering and Psychoacoustics : Matching Signals to the Final Receiver, the Human Auditory Systems"
J. Audio Eng. Soc., Vol.39, No.3, (Mar 1991) pp 115-126.

[5] **M A Gerzon**, "Masking of Coding/Decoding Errors in Audio Data Compression Systems"
Proc. Institute Of Acoustics, Vol.12, Part 8, (Nov 1990) pp 175-182.

[6] **J R Stuart**, "Future Codes : Masking and Data Compression"
Hi-Fi News & Record Review Mag., (June 1991) pp 31-35.

[7] **N S Jayant & P Noll**, *Digital Coding of Waveforms - Principles and Applications to Speech and Video*
publisher Prentice-Hall, Inc., (1984 edt).

[8] **M O J Hawksford & T F Darling**, "Oversampling Analog-to-Digital Conversion for Digital Audio Systems"
J. Audio Eng. Soc., Vol.38, No.12, (Dec 1990) pp 924-943.

[9] **M O J Hawksford**, "Digital Discourse 4"
Hi-Fi News & Record Review, (Aug 1990) pp 39-47.

[10] **P J Naus, E C Dijkmans, E F Stikvoort, A J McKnight, D J Holland & W Bradinal**, "A CMOS stereo 16-bit D/A converter for digital audio"
IEEE J, SC-22, (3), (June 1987).

[11] **J Makhoul**, "Linear Prediction : A Tutorial Review"
Proc. of IEEE, Vol.63, No.4, (Apr 1975) pp 561-580.

[12] **S Singhal & B S Atal**, "Amplitude Optimization and Pitch Prediction in Multipulse Coders"
IEEE Trans. on ASSP, Vol.37, No.3, (Mar 1989) pp 317-327.

5 APPENDIX

Calculation of the Multipulse LPC Bit-Rate and Equivalent Number of Bits/Sample :

(1) Calculation of the sampling rate reduction factor :

$$sampling\ rate\ reduction = \frac{N}{p_\alpha + i_\beta + i_n}$$

where N is the frame size / channel,
p_α is the LPC filter order,
i_n is the number of pulse locations / frame / channel,
i_β is the number of pulse amplitudes / frame / channel.

(2) Calculation of the new sampling rate in kHz :

$$new\ sampling\ rate = \frac{44.1}{sampling\ rate\ reduction} kHz$$

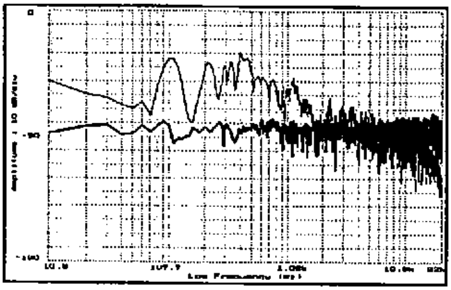
(3) Calculation of the new bit rate in bits/s assuming 16-bits are assigned to encode α_p, β_i, n_i,

new bit-rate = (new sampling rate x 16) bits/s

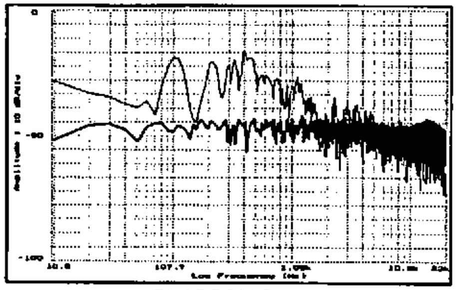
where α_p is the LPC filter coefficients,
β_i is the pulse amplitudes,
n_i is the pulse locations.

(4) Calculation of the equivalent number of bits per sample :

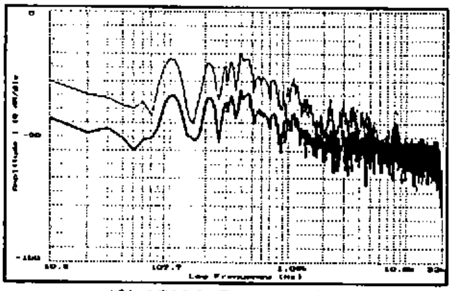
$$equivalent\ no.\ of\ bits/sample = \frac{new\ bit\ rate}{44.1}$$



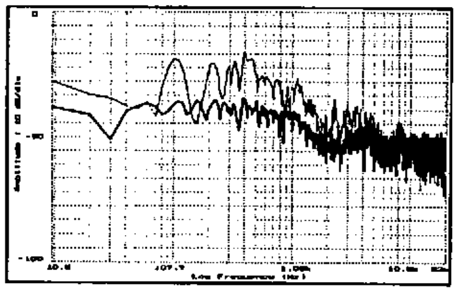
(A) APCM:5-bits



(B) ADPCM:5-bits

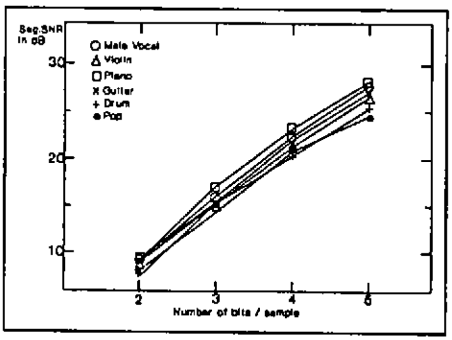


(C) ADM:8xOversampling

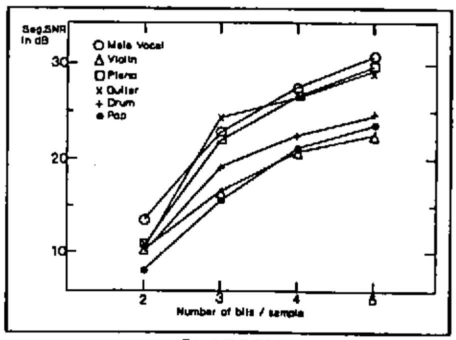


(D) LPC:50-pulses/frame

FIGURES 1 Graphs showing the spectrum of a PCM pop music signal (THIN line) versus the error spectrum (THICK line) of (A) APCM, (B) ADPCM, (C) ADM and (D) LPC coder.

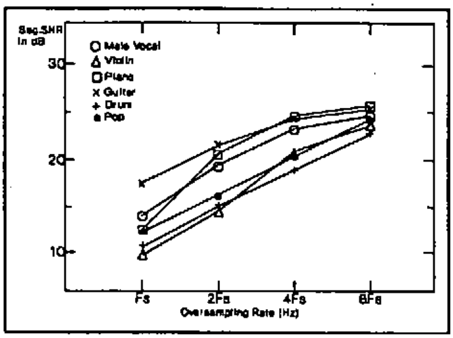


(A) APCM

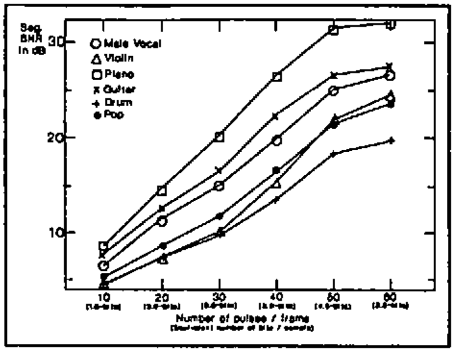


(B) ADPCM

FIGURE 2 ...



(C) ADM



(D) Multi-pulse LPC

FIGURES 2 Graphs showing the coders segmental signal-to-quantization noise ratio (SNR) relationship with six different audio signals for (A) APCM, (B) ADPCM, (C) ADM and (D) Multi-pulse LPC.

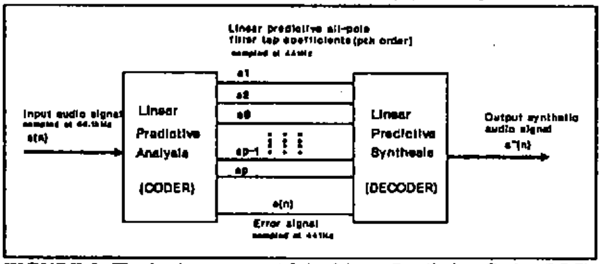


FIGURE 3 The basic structure of the Linear Predictive Coder (LPC).

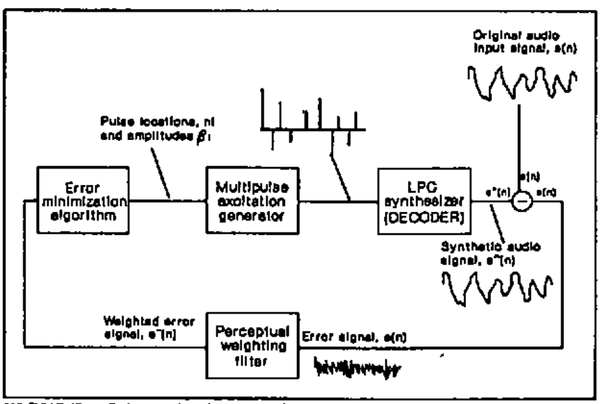
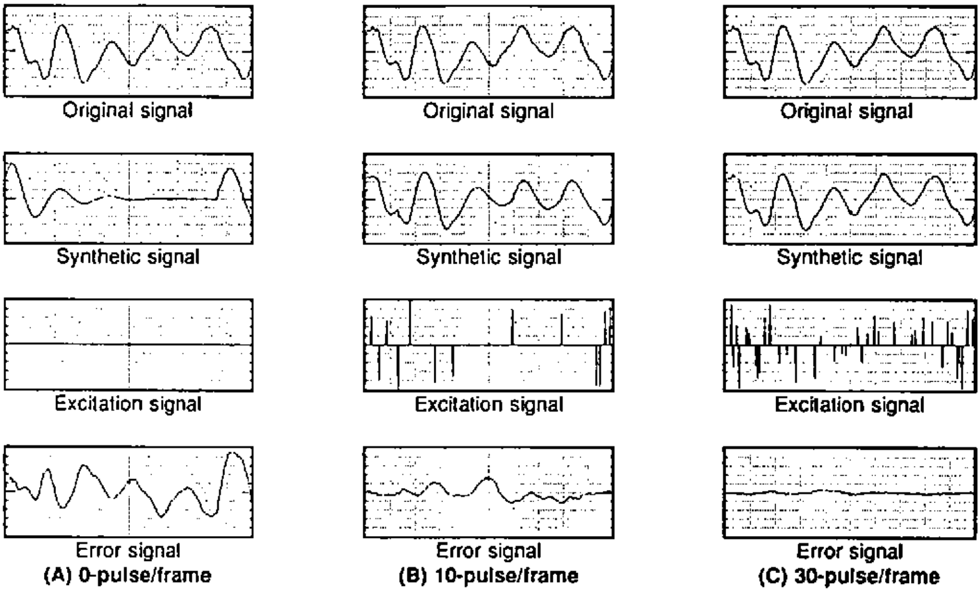


FIGURE 4 Schematic diagram of the multi-pulse excitation LPC-based coder using an analysis-by-synthesis procedure.



FIGURES 5 Waveform analysis of PCM original signal, synthetic signal, excitation signal and error signal for (A) 0-pulse, (B) 10-pulse and (C) 30-pulse (Frame size = 10ms [441 samples] per channel, LPC = 24th order).

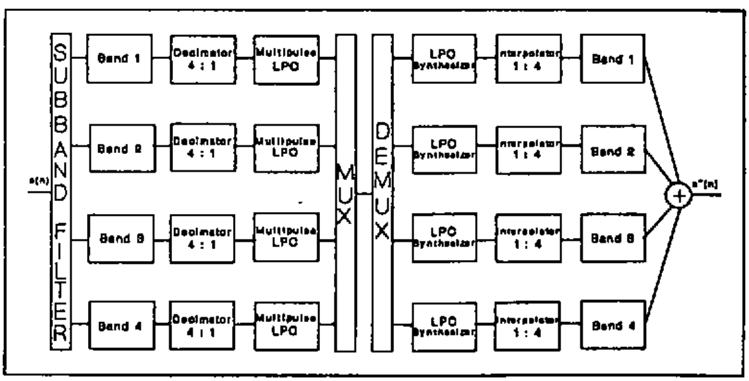


FIGURE 6 Schematic diagram of the proposed Multi-pulse Adaptive Sub-Band Coder (MASC).

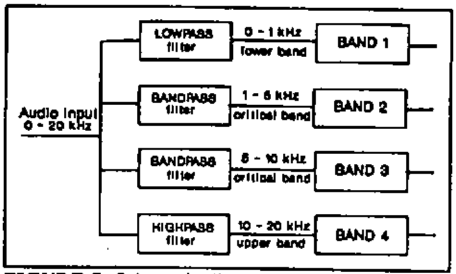


FIGURE 7 Schematic diagram of the 4 x Sub-band filtering design.

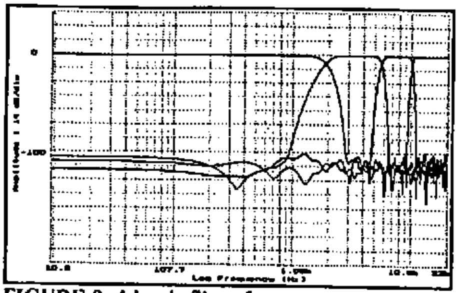


FIGURE 8 4-bands filters frequency response.

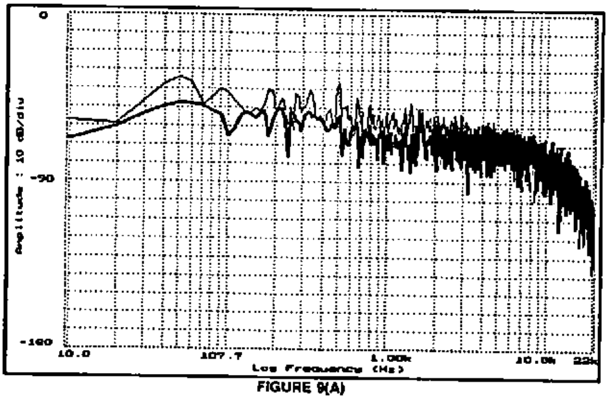
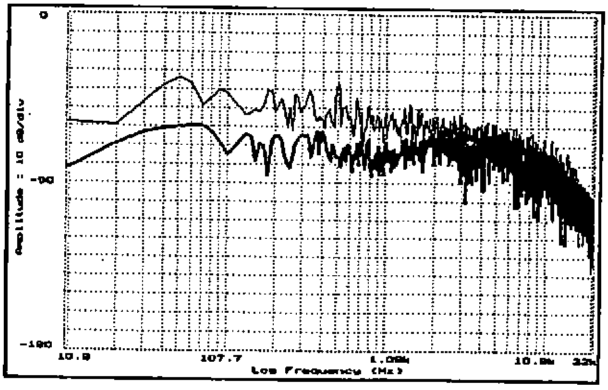
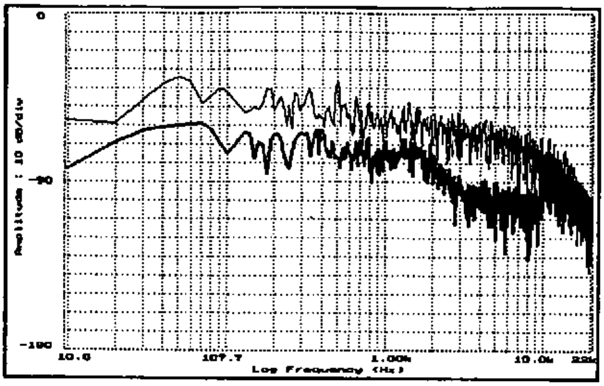


FIGURE 9(A)



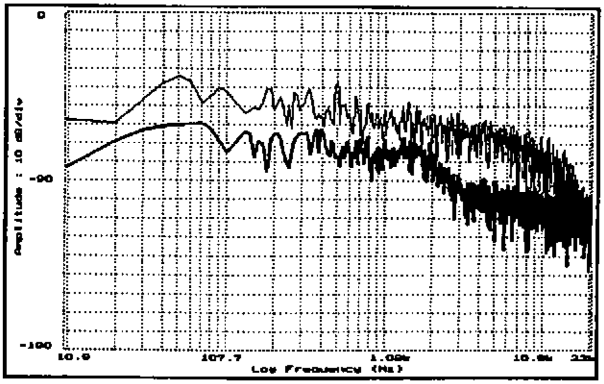
SUB-BAND	Band 1	Band 2	Band 3	Band 4
PULSES	10	10	10	10

FIGURE 9(B)



SUB-BAND	Band 1	Band 2	Band 3	Band 4
PULSES	10	20	20	10

FIGURE 9(C)



SUB-BAND	Band 1	Band 2	Band 3	Band 4
PULSES	10	20	20	20

FIGURE 9(D)

FIGURE 9 Graphs showing the spectrum of a PCM pop music signal (THIN line) versus multi-pulse LPC coding error spectrum (THICK line) at (A) broadband:128kbits/s, (B) sub-band:70.4kbits/s, (C) sub-band:86.4kbits/s and (D) sub-band:94.4kbits/s.