A DIGITAL AUDIO SYSTEM FOR BROADCAST, CABLE, AND SATELLITE DELIVERY MEDIA

C.C. Todd and K.J. Gundry

Dolby Laboratories, San Francisco and London

ABSTRACT

The requirements for a digital audio system to be used for broadcast, cable, and satellite delivery media differ from those for recording in that the major economic consideration is the cost of the playback circuitry. This fact has been utilized in the digital audio system to be described. The system offers high quality sound at a relatively low data rate (on the order of 200-350 k bits/sec) and the option of audio scrambling. No precision components are required in the decoder minimizing cost. The performance, cost and data rate advantages are achieved by placing more sophisticated circuitry in the encoder. Applications being pursued include DBS, cable, pay-TV, and terrestrial broadcast systems.

INTRODUCTION

The system described in this paper (1) arose from Dolby Laboratories' continuing work on exploiting digital transmission and recording techniques without the high costs inherent in PCM. Previous papers (2) (3) concerned a proposed use of delta modulation for recording television sound on magnetic tape; a consumer VTR incorporating that system would contain both an ADC and a DAC (although much of the circuitry could be in common). In broadcasting only the DAC appears in the consumer's home. This paper describes a digital audio broadcasting system in which the DAC has been further simplified, at the expense of greater complexity in the (professional) ADC. The values assigned in the paper to various circuit constants reflect the probable first application, direct broadcasting of television sound via satellites, (525 line, 60 field/sec) but it will be apparent that the principles lend themselves to many other media.

Although much of the discussion concerns two-level differential quantizing (delta modulation) the noise reduction techniques could equally be applied to multi-level differential systems. However, this would not satisfy one of the design aims, elimination of the multi-level DAC, a component necessitating high precision in manufacturing.

PERCEIVED OBJECTIVES

We set ourselves the following goals:

A. We demanded a subjectively transparent system for delivery of the highest quality audio from the broadcast center to the consumer. The system should not introduce audible degradation to audio program material of the highest quality (not only the highest quality which is available today but the highest quality which can be expected to be available in the future).

B. The system should have a high tolerance of errors, such that only a modest degradation of audio quality shall be perceived when "worst case" error conditions occur.

C. The system should be economical. The receiving equipment should be very low in cost. The system should be efficient in usage of channel capacity since then more excess capacity will exist for flexibility to add additional channels, or revenue generating services, or more bandwidth will be available to the video signal.

D. The system should be practical in operation. The transmission equipment should not require any special attention (such as very accurate level adjustment to prevent clipping) which might exceed the capabilities of the broadcast personnel, or require the use of noncomplementary signal processing.

POSSIBLE SOLUTIONS

Considering these goals, we compared various digital encoding schemes.

A. High bit-rate linear PCM with sophisticated error corection.

This brute force approach can satisfy objectives A, B and D easily. However it falls short on objective C. It requires bit-rates of 700 kbit/s or more per audio channel, and is inherently expensive.

B. Efficient PCM with modest error correction.

Digital companding may be used to reduce the bit rate of solution A. Digital compandors resemble analog compandors in that the level of quantizing noise rises as the signal level rises; that is, digital compandors are fundamentally susceptible to noise modulation (defined as a modulation of the perceived background noise level by the audio signal). If the transmitted code has at least 10-bit resolution in the presence of full scale signals, and high frequency pre- and de-emphasis are used, then acceptable performance can be achieved. The resulting loss of high frequency headroom is acceptable with audio program material. The error correction system may be simplified if modest degradations in audio quality at high error rates are accepted and error concealment is employed to lessen the requirements for error correction. Bit rates can then be reduced to perhaps 350 kbit/s.

However the cost saving of this approach is only modest (if any) compared with solution A because the precision of the required DAC has not been reduced. The costs are dominated by the requirements for precise 14 to 16-bit D-A converters, sharp cut-off low pass filters, and the error correction circuitry.

C. Adaptive Delta-Modulation

Delta modulation has some significant virtues. The circuitry does not require any high precision components, and can be manufactured very economically with today's technology. Since all bits have essentially equal significance, isolated bit errors have a minor audible effect. One can consider operation without an error correction system, yielding a saving in bit rate and cost. However, linear delta modulation at the bit-rates under consideration (a few hundred kbit/s) has an inadequate dynamic range for high quality audio. We are therefore led to consider adaptive delta modulation (ADM).

Adaptive Delta-Modulation is a companded system so noise modulation must be considered. In contrast to companded PCM, noise modulation in an ADM system is caused not by high amplitude signals but by high slope signals. Noise modulation is worst in the presence of high frequency signals but the fact that high frequency signals effectively mask noise makes this characteristic acceptable and even preferable to that of companded PCM where high amplitude low frequency signals will produce noise modulation which may not be masked by the signals.

The usual technique for ADM involves coding into a single bit-stream both the audio information and the step-size or scaling information, so that the adaption of the delta modulator is necessarily output controlled. The significance of a bit then varies in accordance with the characteristics of the adaption algorithm, leading to a "gain blipping" effect on a small percentage of errors (which have hit "critical" gain control bits); this is the most noticeable degradation caused by errors on such a system. The effect can be reduced in magnitude by making the control signal move more sluggishly at the expense of poorer transient response. The adaption characteristic is therefore a compromise between speed of response (necessary for handling transient signals) and tolerance of errors; that is, objectives A and B above cannot be met together.

A judicious choice of fixed pre- and de-emphasis characteristic can provide a good compromise between audible noise modulation with mid frequency signals and high frequency handling capability, but is unsatisfactory when program material contains predominantly high frequency energy. The problem is not so much that the signals might overload the system, but that in this situation the de-emphasis curve in the decoder is more accurately described as a low frequency emphasis characteristic which actually increases low frequency noise (see section 7 below). Output controlled ADM with fixed emphasis may satisfy objective C, but is unlikely to meet objectives A and B.

OUTLINE OF NEW SYSTEM

During our work on digital audio for video tape we became very well acquainted with the promises and pitfalls of delta modulation.

A fundamental aspect of this new approach is that we have significantly increased the complexity and cost of the encoder (of which very few are required) in order to lower the cost of a decoder and to remove the performance limitations of a simple ADM system.

The "gain blipping" effect of a conventional ADM system has been substantially eliminated by extracting the step-size control information and then generating from it a separate low data rate bit stream. A simple algorithm is employed to convert this bit stream into a control signal of limited bandwidth and with this algorithm all bits have equal weight. It is thus impossible for an occasional bit error to cause much of a gain variation.

A low data rate control signal suggests a sluggish response which would yield poor transient performance. We have avoided degrading transient response by employing a delay line in the encoder. This technique allows a sluggish control signal to begin to rise in anticipation of an oncoming audio transient. The conventional tradeoff is not necessary and perfect transient performance is achievable with no additional cost in the decoder.

The performance compromise inherent in fixed pre-and de-emphasis is removed by the use of a variation on our proprietary "sliding band" pre-and de-emphasis. This powerful technique gives a larger improvement in noise modulation than fixed emphasis yet does not incur a penalty of reduced high frequency headroom or low frequency noise emphasis in the presence of predominantly high frequency program material.

The method of controlling the sliding band is similar to that used for the step-size. A circuit in the encoder analyzes the spectral distribution of the program material to determine the optimum placement of the sliding band. After a delay this emphasis is applied to the audio, and the control signal is encoded into a separate bit stream which is handled identically to the step-size bit stream. Again, because of the delay line, perfect dynamics are achieved by the encoding method. The decoder complexity is not affected.

Before explaining in greater detail, it is necessary to cover some theoretical points.

USE OF LIMITED BANDWIDTH CONTROL INFORMATION

In adaptive delta modulation, the step-size or unit of quantization is made variable, increasing with increasing slope of the input audio signal. The operation of the adaptive modulator is equivalent to sampling and quantizing an audio waveform which has been multiplied by a further waveform representing the variation of step-size.

When one waveform is multiplied by another, as in amplitude modulation, the output signal has an extended spectrum; in this case, each spectral line of the modulating (step-size) waveform fm adds a pair of modulation side-bands to each spectral line fa of the input audio, spaced from fa by ±fm.

Similarly the action of the adaptive demodulator is to introduce side-bands. In a perfect adaptive system, the modulation and demodulation products will have exactly equal amplitudes but opposite polarities, and will therefore cancel leaving only the desired audio. This discussion is true whether the adaptation occurs "instantaneously" that is, the step-size can change from one value to a distinctly different one between two adjacent sampling periods, or whether it occurs smoothly and is continuously variable.

In a real-world system, the encoder and decoder will not track perfectly because of component tolerances and/or transmission bit errors. In an instantaneous system, discrepancies between the encoding and decoding step-sizes leave modulation products spreading across the whole audio spectrum; if the noise and distortion of the overall system are not to be degraded audibly, the step-sizes must be defined with an accuracy comparable with the minimum resolution of the quantization.

In a continuously variable system, the modulation products resulting from mistracking are the side-bands mentioned above, attenuated by partial cancellation. The lower the bandwidth of the step-size control waveform (the smaller fm), the more tightly the modulation products are confined to the immediate vicinity of each audio spectral component.

The masking properties of any particular audio frequency are greatest near that frequency. For example, a 1 kHz signal makes low level noise and distortion components within a few hundred Hz of 1 kHz inaudible, but components at the extremes of the audio spectrum remain audible. It is this property of human hearing that makes it possible to design noise reduction systems without audible noise modulation.

Hence the narrower the bandwidth of the step-size control, the less audible will be the modulation products of mistracking, or the greater amount of mistracking can be tolerated. With the control bandwidth employed in our new system, approximately 50 Hz, mistracking of several percent remains inaudible on real program material, and component tolerances can be greatly relaxed. Furthermore error rates up to perhaps 10^{-2} in the control bit-stream can be accepted without correction.

OPTIMIZED ADAPTIVE DELTA MODULATION

The noise and distortion emerging from a complete ADM codec depends on (among other things) the nature of the audio input signal and the step-size; both these factors are varying all the time.

Consider a codec receiving and reproducing the simplest audio waveform, a constant amplitude sine-wave. As a function of step-size, the output noise and distortion will vary as shown qualitatively in figure 1.

In the region labelled a, the step-size is unnecessarily large, giving excessive quantizing noise. In region b the step-size is too small and the system is therefore in slope overload, giving high distortion. There is an optimum value of step-size for the particular input conditions, labelled c.

For each short time segment of real audio there is a curve like figure 1, and there is a best step-size. With discrete instantaneous adaptation in which the step-size can only adopt one of a limited number of values, it is clearly impossible to operate at the best values all the time, since they will inevitably lie between the available values.

With continuously variable adaptation it is possible to operate very near the best values. In a system employing limited bandwidth control information, the relevant time segment is a window related to the rise-time of the step-size control signal, in our case around 10 ms. From examination of the audio within a moving window of duration 10 or 15 ms, an optimum step-size can be generated which minimizes the noise and distortion from the codec.

Note that a conventional output controlled ADM system, in which the bit-stream carries two pieces of information, the audio and the step-size, cannot achieve optimum step-size, but remains in region a most of the time, moving into region b on signal transients. Also, of course, adequate response time for signal transients indicates that the bandwidth of the step-size adaptation must be wide (many kHz), implying a need for much greater precision.

VARIABLE PRE- AND DE-EMPHASIS

In an audio system whose noise is independent of the audio (for example, an amplifier with thermal noise or an FM broadcast system), the effects of fixed pre-and de-emphasis are easy to understand and to calculate.

The choice of emphasis characteristic is usually a compromise between noise reduction and overload characteristic, and in some applications noise reduction effect is offset by the need to reduce program level to the extent that little improvement in unweighted signal-to-noise ratio is observed. However the change in the spectrum of the noise may be audibly valuable.

In an optimized ADM system, the noise varies with the signal. At any fixed input frequency, the noise is directly proportional to the signal amplitude (that is, the instantaneous signal-to-noise ratio is constant), and at any fixed input amplitude the noise is directly proportional to signal frequency. Furthermore, unlike PCM, there is no clear system-defined maximum signal level which can be transmitted, although practical instrumentation may impose one.

The effects of fixed high frequency pre- and de-emphasis on such a codec are quite different. When the predominant audio input spectral components are at low or middle frequencies, that is in the unboosted area, emphasis does not change the demand on step-size in the ADC or DAC, the basic noise output of the demodulator is unchanged by emphasis, and hence subsequent de-emphasis reduces high frequency noise.

However, when a predominant audio input spectral component is within the area of frequencies boosted by emphasis, an increase step-size is demanded, with the result of increased noise output from the demodulator. The de-emphasis then does several things:

i) It restores the audio component to its correct (unboosted) level.

ii) It lowers the noise in the region of that component, but only back to the level it would have had without pre- and de-emphasis.

iii) It lowers the noise at frequencies above the audio component, but starting from the increased level.

iv) It has no effect on the low frequency noise, which therefore remains at its increased level.

Hence for high frequency signals, the effect of fixed pre- and de-emphasis is not primarily to degrade headroom or give distortion, but to <u>increase</u> low frequency noise, that is, the noise which is <u>least likely</u> to be masked by high frequency signals.

Consideration of the noise levels and spectra delivered by an optimized ADM codec operating at 200 or 300 kbit/s shows that there are three (overlapping) regimes to be considered if satisfactory noise shaping is to be employed.

A. When the predominant audio spectral components lie below roughly 500 Hz, a large high frequency pre- and de-emphasis will reduce noise sufficiently that audible noise modulation will not occur. An example of a practical emphasis characteristic is a 20 dB shelf starting at 800 Hz (see curve 1 on figure 2).

B. As the predominant spectral component is increased up to 2 or 3 kHz, it is necessary to slide the emphasis upwards in frequency so as to retain high frequency noise reduction <u>relative to the spectral component</u> (curves 2 and 3). <u>Low frequency noise is not yet an audible problem.</u>

C. When the predominant spectral component is above about 3 kHz, noise both at low and at very high frequencies must be reduced. An emphasis curve with a dip at the predominant component will reduce the step-size and hence the broad-band noise emerging from the codec, while the subsequent complementary de-emphasis peak will pick out the wanted signal component, while attenuating high frequency noise, and retaining the reduced low frequency noise level which resulted from the smaller step-size. For example, if the predominant signal lies at 6 kHz, curve 5 is a desirable emphasis characteristic.

This explanation has assumed that the predominant components of an audio signal at a particular moment are concentrated in a narrow region of the spectrum; such a signal is in fact the most critical case. When the spectral components are more distributed, their masking properties cover more of the noise, and the emphasis shape is less critical.

Thus a variable emphasis circuit giving a family of response curves of the form shown in figure 2 preceding the modulator, with the complementary de-emphasis following the delta demodulator, will provide efficient subjective noise reduction under all signal conditions, provided that circuitry can be designed to analyse the input audio and to give suitable instructions to the emphasis and de-emphasis.

<u>PRACTICAL DETAILS OF THE DECODER</u>

Figure 4 illustrates the decoder required in the consumer's home. Each audio channel of decoding receives three data bit-streams.

The audio data is at a moderately high rate, 200 kbit/s or more. The precise rate depends on the application; for television it is convenient to operate with an integral number of bits per horizontal picture line. The two control bit-streams are at much lower rates; for television a practical rate is half the television horizontal frequency.

<u>Basic Adaptive Delta Demodulator</u>

In order to achieve the required dynamic range from adaptive delta modulation, it is necessary to be able to adapt the step-size or unit of quantization over a range approaching 50 dB. The basic demodulator takes a step-size voltage Vss (or an equivalent current) which can vary over this range and switches it with one polarity or the reverse in accordance with the audio data to a leaky integrator. The leak time-constant corresponds to a few hundred Hz; below this frequency the system is strictly not delta but delta-sigma modulation.

Step-Size Decoder

The step-size control bit-stream contains the logarithm of the required step-size coded as delta-sigma modulation. It is therefore decoded by passage via a low-pass filter, which determines the bandwidth (and hence rise-time) and ripple of the step-size voltage, and an exponentiation or anti-logarithm. circuit (for example, a bipolar transistor, which inherently has the appropriate characteristic). If the normalized mean level of the bit-stream (or the duty-cycle measured over the rise-time of the low pass filter) is written as x, then

$$V_{ss} = V_o \exp kx$$

where Vo is a constant suitable for the particular instrumentation, and $k = 10 \ln 2$

This definition gives an increase of 6 dB in step-size for every increase of 0.1 in x. Since by definition x is confined to the range 0 to 1, the resultant maximum possible range of Vss is 60 dB. Of this, about 50 dB is useful.

The transmission of step-size information in logarithmic form reduces the dynamic range conveyed in the bit-stream, in this case from about 50 dB to about 9 times, or 19 dB, and spreads the effect of bit errors more uniformly across the dynamic range. Since Vss is confined by the low-pass filter to a bandwidth of about 50 Hz, bit errors lead to slow random amplitude modulation of the output audio. The audible disturbance produced by errors in the control bit-stream is quite negligible compared with the effect of similar error rates in the audio data. The control system is so robust that uncorrected error rates up to 10^{-2} or so produce nearly imperceptible disturbance of music or speech.

Sliding-Band De-Emphasis

The requirements have been discussed in section 7; figure 3 shows a representative set of de-emphasis curves, complementary to the emphasis of figure 2.

There are obviously many ways of synthesizing responses of this nature. Figure 4 shows one practical technique; here the cut-off frequency of the high-pass filter formed by the capacitor and variable resistor should be directly proportional to the control signal decoded from the sliding-band control bit-stream.

Those familiar with Dolby noise reduction systems (such as A- or B-type) will recognize the dual-path configuration in which a main path with fixed characteristics is paralleled either with feedforward or feedback by further paths having variable characteristics. However in noise reduction the further paths constitute compressors, while in this new system the further path (the variable high-pass filter of figure 4) is ultimately controlled by the spectrum of the incoming audio; there is no systematic compression or expansion of the dynamic range.

Emphasis Control Decoder

The emphasis control decoder is substantially identical with the step-size decoder. The emphasis control bit-stream contains the logarithm of the cut-off frequency of the variable high-pass filter in the de-emphasis circuit; the virtues and benefits of a logarithm. function here are similar to those in the step-size control. The cut-off frequency follows the relationship

$$f_1 = f_o \exp ky$$

where fo is a constant determining the scaling, $k = 10 \ln 2$, and y is the normalized mean level of the bit-stream

This definition results in a movement of one octave for every increase of 0.1 in y. The control of emphasis in the presence of errors is if anything more robust than that of step-size.

Output Filter (not shown on figure 4)

Since in delta modulation the sampling frequency is vastly greater than the minimum required by information theory, there is little probability of aliasing components in the output. Non-audio spectral components in the output are at frequencies well above the audio band, and hence only an elementary two or three pole low-pass is necessary.

REQUIREMENTS OF THE ENCODER

As will be clear from the block diagram in figure 5, the encoder is very much more elaborate and expensive than the decoder; this is not a serious objection since it does not need to be mass produced for the consumer market.

The basic requirement, of course, is to deliver bit-streams which can be interpreted correctly by the decoder.

The emphasis control block analyzes the spectrum of the input audio to determine which of the family of available emphasis and de-emphasis curves will minimize the audibility of codec noise in the presence of that signal spectrum. This information is coded into the low-bit rate emphasis control data stream. The conversion of this bit-stream back to a signal suitable for operating on the variable filters of the emphasis and de-emphasis includes limitation of the bandwidth to about 50 Hz, corresponding to a rise-time of about 10 ms. Hence the input audio is delayed by this time before entering the variable emphasis circuit.

After emphasis, the audio is passed to the step-size detection block which measures the slope of the signal to determine the optimum step-size (point c on the curve in figure 1). The logarithm of this value is coded into the step-size control data. Again conversion back to actual step-size voltage Vss restricts the rise-time, and so the emphasized audio is delayed by about 10 ms before application to the main encoding delta modulator.

Since the audio is subject to this further delay after emphasis, the emphasis control bit-stream needs delaying also so that all the data will arrive synchronously at the decoder.

ERROR CONCEALMENT

Our previous work (1) illustrated the nature of the difference between a bit error reproduced by a PCM or DM system. In the PCM case the reproduced error takes the form of a narrow impulse, the height of which depends on which bit is hit (being very large for an MSB hit). In the case of delta modulation, the error takes the form of a small step of significant length (decaying only because the integrator has a "leak" with a 0.5 msec time constant); small because there is no possibility of hitting an MSB (there isn't one - all bits are equal!). If the error rate is low (on the order of 10^{-5}) a delta modulation system is quite usable without any error correction or concealment, while a PCM system is not.

PCM systems intended for operation at moderate to high error rates (10^{-4} to 10^{3}) must (to be usable) carry some overhead for error concealment. No attempt is made to correctly reproduce the audio, but using a small amount of redundant data, bad words are identified and an interpolated value substituted. A very substantial audible improvement can be realized with a modest amount of data overhead and a reasonable amount of additional circuitry (storage of a previous sample, addition of the next sample and shift right for divide by two). The result of concealment is to greatly reduce the amplitude of the impulse which is added to the reproduced audio.

Delta modulation systems can benefit from a somewhat analogous error concealment scheme which yields a similar result with simpler circuitry. The scheme involves detection of the approximate location and polarity of the error, and the creation of an error of opposite polarity in nearly the same location. This has the effect of terminating the duration of the step the error has created yielding an impulse similar to the concealed PCM impulse.

The implementation of the scheme involves blocking the data into N bit blocks and performing a modulo 4 summation of the number of data 1's in the block. This two bit result (0, 1, 2, 3, 0, 1, 2, 3 . . .) is transmitted along with the data block as redundant information. If a single bit error occurs in the data block (for moderate error rates nearly all errors will be single errors) the number of data 1's in the received block will differ from that sent. If the same modulo 4 summation is performed in the decoder and compared to the result received from the transmitter, the presence and polarity of a single bit error is detected. If an error is detected, the concealment involves creating an error of opposite polarity in the block. Since we have only located the error to a particular block and we want our artificial error to be located as close as possible to the real error, we insert the artificial error into the center of the block. The distance between the real and artificial errors will then be between 0 and N/2 bits, with an average distance of N/4 bits. The step introduced into the audio will be (on average) this length. Note that some errors ($2/(N+2)$) will hit the concealment data and the system may misconceal. This possibility can be virtually eliminated by carrying along a 3rd bit which is a parity on the other two, but the audible improvement doesn't really justify the added overhead unless the block size is quite small ($2/(N+2)$ becoming large).

By way of example, suppose we block the data into 32 bit blocks and send 2 redundant bits for each block for an overhead of about 6%. For approximately 32/34 ths of all errors (94%) we correctly receive the concealment data and can conceal the error, terminating it to an average length of N/4 or 8 bit periods. For a system operating at a 250 kHz data rate, the step would have a length of 32 μsec (about the same length as an impulse in a PCM system sampled at 32 kHz). In practical sytems the data is often organized in blocks of some size (in a video application a block of data may be sent every horizontal scan interval), so these same blocks are used to implement the concealment.

CONCLUSION

This paper has described a digital audio system particularly suited to broadcasting, cable, and satellite delivery. It offers a bandwidth in excess of 15 kHz, a dynamic range potentially around 100 dB (although initial embodiments may be limited to perhaps 85 dB because of the limits of present IC technology and of the available means of producing 10 ms delays), audio scrambling, and a level of program-modulated noise better than most of the digitally-companded PCM systems proposed or in use. All this with a total bit-rate potentially less than any existing PCM system, and with a decoder having a fraction of the cost.

REFERENCE

(1) This paper, in substantially its present form, was presented at the 75th Convention of the Audio Engineering Society, Paris, March 1984 (preprint 2071 (C7)).

(2) K.J. Gundry, D.P. Robinson and C.C. Todd, "Recent developments in digital audio techniques", presented at 73rd Convention of the Audio Engineering Society, Eindhoven, March 1983 (preprint 1956).

(3) S. Forshay, K. Gundry, D. Robinson, C. Todd, "Audio Noise Reduction in Cable Systems," presented at 32nd Annual National Cable Television Association Convention, Houston, June 1983.

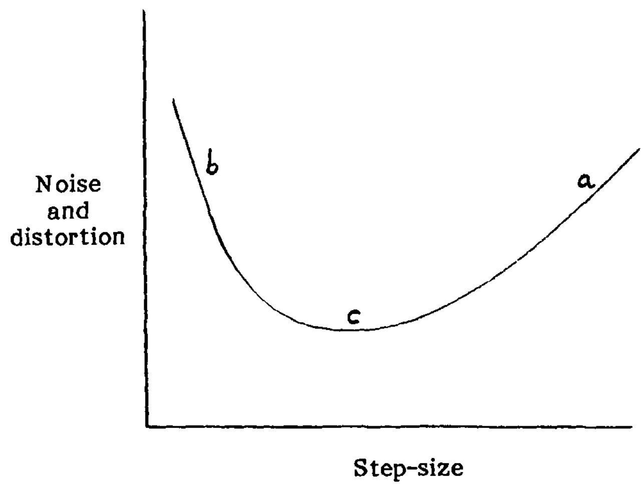

Figure 1

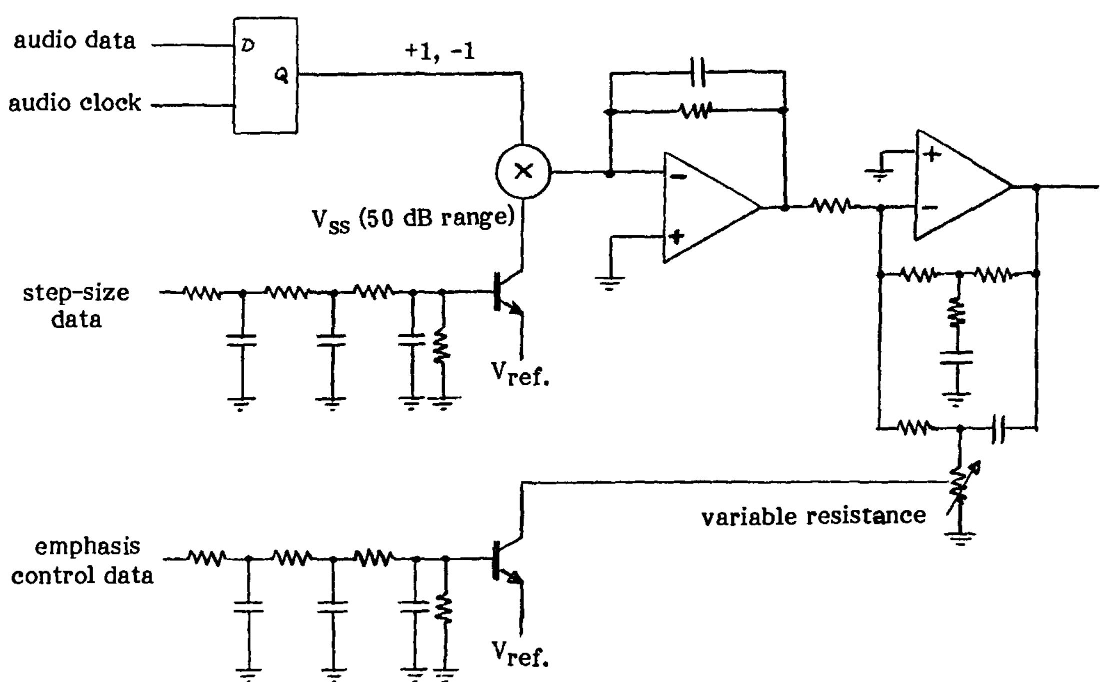

Figure 4 Consumer decoder

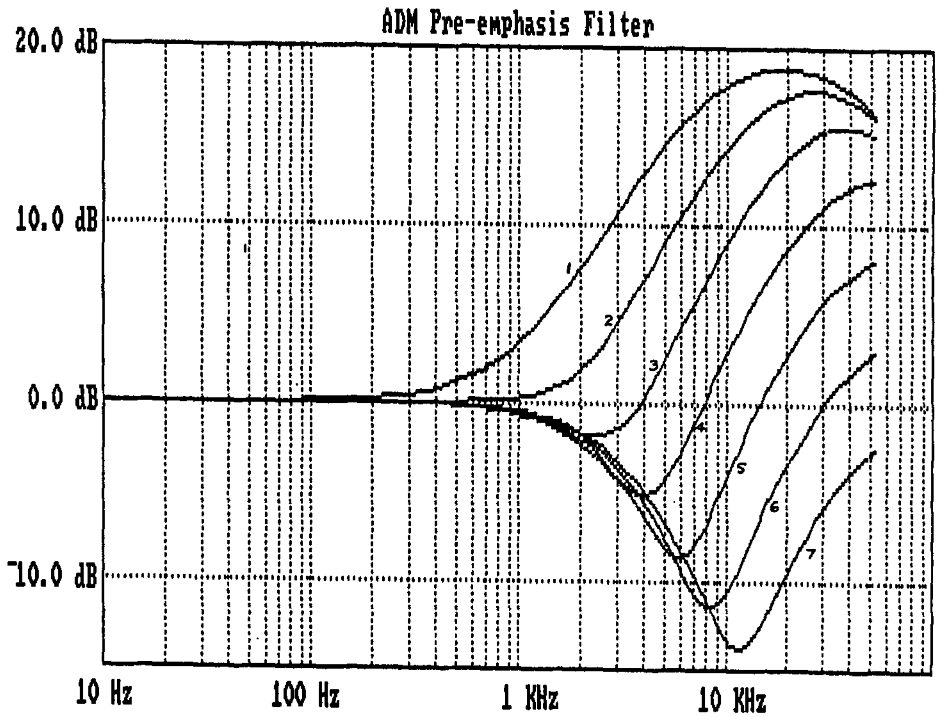

Figure 2

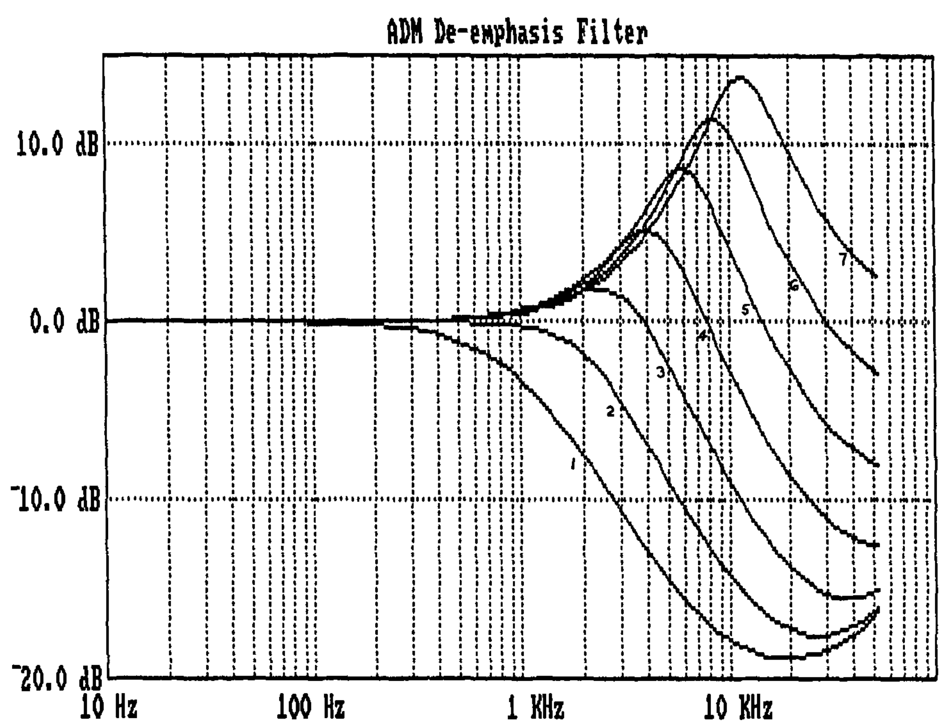

Figure 3

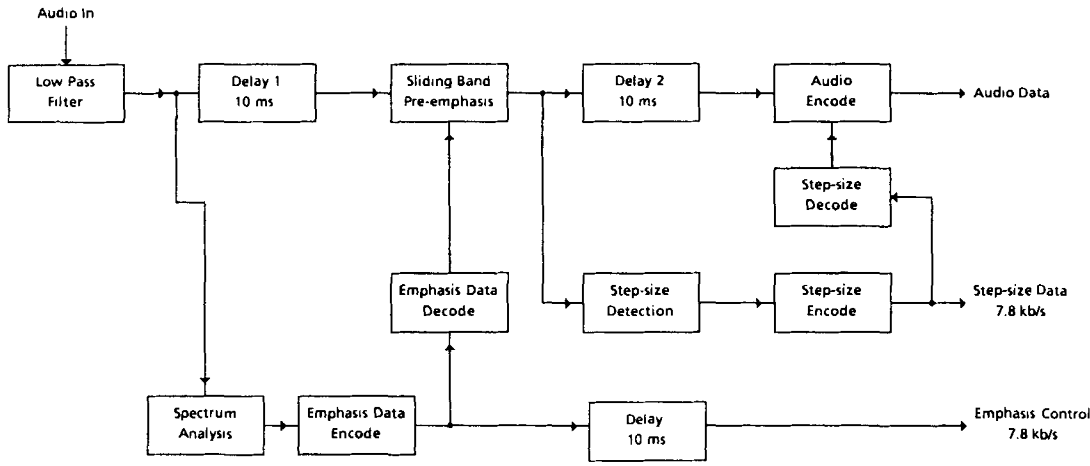

Figure 5. Professional Audio Encoder. Block Diagram