음성파형 부호화에서의 잡음 SPECTRUM 변형에 관한 연구

(Noise Spectral Shaping in Speech Waveform Coding)

* 이 황 수 (H. S. Lee)
* 은 종 관 (C. K. Un)

ABSTRACT

The performances of speech waveform coders that incorporate noise spectral shaping have been studied with real speech. The coders studied were adaptive pulse code modulation (APCM), adaptive differential PCM (ADPCM) and adaptive delta modulation (ADM). Two types of noise spectral shaping methods are used to shape the spectral components of quantization noise at the decoder output. For APCM and ADPCM, the noise feedback filter is designed to minimize the C-message weighted quantization noise power. As for ADM, the noise feedback filter is used to move some portion of in-band noise outside the signal band. The performance criteria used for these coders are frequency-weighted signal-to-quantization noise ratio (FWSQNR) and segmental FWSQNR ($FWSQNR_{SEG}$) for APCM and ADPCM, and $SQNR_{SEG}$ for ADM. A method to disign a stable noise feedback filter is proposed. Simulation results with real speech input show that the performance improvements of these waveform coders with spectral shaping are about 0.5 to 3 dB over the system without noise shaping. Although this improvement in SQNR is relatively small, the waveform coders with noise spectral shaping yield, according to our informal listening test, subjectively more pleasing and intelligible sound than the conventional waveform coders without noise spectral shaping.

要　　約

본 논문에서는 잡음 spectrum 변형 기능을 가진 APCM, ADPCM 및 ADM 음성 부호기의 성능에 관해서 연구하였다. 잡음 spectrum 변형방식은 두가지를 고려할 수 있는데, APCM과 ADPCM 에서는 C-message weighting된 양자화 잡음을 최소화하는 noise feedback filter를 이용하는 방법을 채택하고, ADM에서는 in-band의 잡음의 일부를 신호대역의 밖으로 옮기는 방법을 사용하였다. APCM과 ADPCM 부호기의 성능을 측정하는데는 주파수가 weighting이 된 신호대 잡음비(FWSQNR)와 segment된 FWSQNR($FWSQNR_{SEG}$)를 사용하였다. 실제음성을 사용한 simulation 결과에 의하면 잡음 spectrum 변형기능을 가진 부호기가 없는 것보다 0.5내지 3dB 가량 좋은 것으로 나타났다. 이러한 개선은 양적으로 비교적 적은 것이 사실이지만 실제로 음성을 들어보면 음질이 현저히 좋아짐을 알 수 있었다.

*한국과학기술원 교수

I. INTRODUCTION

It has long been recognized that the perceptual effects of noise vary with its frequency spectrum. Since the human auditory system acts like a frequency-selective filter bank, the human perception of noise is frequency-variant [1]. Recent experimental results indicate that the subjective loudness of quantization noise can be reduced by spectrally shaping quantization noise in relation to input speech spectrum [2]. One may note that most of waveform coders have been designed to minimize the unweighted mean-squared error between the system input s_n and the output $\hat{s}_n$. These coders have not been designed to minimize the perceptual effect of noise.

One typical waveform coder structure designed to provide the desired spectral shaping of noise is the noise feedback coder (NFC) proposed by Noll [3]. The NFC is a differential encoding system that has a configuration similar to differential pulse code modulation (DPCM). But, unlike DPCM in which a feedback circuit is used to predict the input signal, the goal of the NFC is to produce a perceptually pleasing output by shaping the quantization noise spectrum. To accomplish this objective, the quantization noise is fed back to the quantizer input through a filter whose parameter values are adjusted such that the desired subjective effects can be achieved.

The use of a noise feedback circuit for reduction of quantization noise was first attempted by Spang and Schultheiss [4]. They could reduce noise power considerably by a slight increase in sampling rate. An increse of 25 percent in sampling rate provided a 95 percent decrease in noise power. This could be accomplished by redistributing the error power over the frequency band in a manner determined by the feedback filter that reduces the noise in the signal bank by adding more noise to the frequency region outside the signal band.

The approach used by Kimme and Kuo [5] is different from that of Spang and Schultheiss in that the sampling rate of the coder was the same as the Nyquist rate. They used a noise feedback loop around the quantizer for coding of narrow band television signals so as to minimize the subjective effects of quantization noise. In their work, they minimized the frequency-weighted integral of quantization noise power spectral density.

In coding of speech, noise spectral shaping has so far been applied mainly to adaptive predictive coding (APC) systems. Atal and Schroeder found that adaptive noise shaping improved the subjective quality of the APC system with an adaptive pitch predictor [6], [7]. Makhoul and Berouti independently reported similar results for an APC system with entropy coding [8], [9]. The bit rates of these coders were about 16 kbits/s.

In addition, Bastian showed that adaptive noise spectral shaping improved the subjective performance of adaptive DPCM (ADPCM) with a fixed predictor at the bit rates of 24 kbits/s and 32 kbits/s [10]. Copperi et al. applied noise shaping to an instantaneously companding ADM to reduce the quantization noise power in the baseband [11].

As a result of those studies, the effectiveness of noise spectral shaping in speech coding is fairly well-known. However, the exact performance improvements of various waveform coders with noise spectral shaping over those without it have not been reported. Furthermore, a detailed study on the design of noise shaping filters specifically for waveform coders has not been done. The major objective of this paper is to investigate how one can effectively design and incorporate a noise shaping filter in the conventional waveform coder such as adaptive PCM (APCM), ADPCM and ADM, and then to study how much one can improve the performance of these coders by using the noise feedback filter. The bit rates of interest are 4 bits/sample (i.e., 32 kbits/s) for APCM and ADPCM, and 16, 24 and 32 kbits/s for ADM. The performance criteria used are the frequency-weighted signal-to-quantization noise ratio (FWSQNR) and the frequency-weighted segmental SQNR ($FWSQNR_{SEG}$) for APCM and ADPCM, and $SQNR_{SEG}$ for ADM, which are known to be subjectively correlated.

The basic objective of using a noise shaping filter is to distribute the noise power in relation to the input speech spectrum so that the noise is masked by speech. In previous works on APC and ADPCM, the coefficients of a noise shaping filter were updated at every short time interval according to input speech variation. This type of adaptive noise shaping in APCM and ADPCM requires real time computations of linear predictive coding (LPC) coefficients for updating coefficients of the noise feedback filter in every frame of speech. This means that the incorporation of adaptive noise shaping in APCM and ADPCM may be impractical because of the coder complexity. Thus, the use of a fixed noise shaping filter that improves the coder performance, yet requires far less coder complexity would be practically preferrable. Therefore, in this paper we study the performances of the waveform coders with fixed noise spectral shaping.

If one uses a fixed noise shaping filter, it is important to choose a suitable subjective fidelity criterion in the frequency domain since the noise shaping filter should be designed such that the subjective quality measure is maximized. There are various measures that use frequency weighting. Some of them are derived from the articulation index of speech intelligibility [12], subjective loudness [13], and the quality measure of telephone speech channels (i.e., C-message weighting [14], [15]). In addition, Miller has found that noise in the range of 1 to 3 kHz is more destructive in intelligibility than that at lower frequencies [16]. Thus, to obtain the frequency-weighted noise power, we use a C-message weighting function which emphasizes high frequency noise relative to low frequency noise.

As for noise shaping for ADM signal, the situation is somewhat different from that of APCM and ADPCM. The major difference is that the sampling rate of ADM is several times the Nyquist rate normally used in PCM or DPCM. Thus, we use a simple high-pass filter as a noise feedback filter in this case.

Following this introduction, general aspects of noise spectral shaping are discussed in Section II. In Section III we consider the design methods of a noise shaping filter. In Section IV coder structures of APCM with noise spectral shaping (APCM-NS), ADPCM-NS and ADM-NS are presented. Simulation results and discussion follow in Section V. Finally, conclusions are made in Section VI.

II. GENERAL CONSIDERATION OF NOISE SPECTRAL SHAPING

In general, two types of noise spectral shaping schemes can be applied to a waveform coder. Fig. 1 shows one type of noise spectral shaping which incorporates preemphasis and deemphasis. In this figure, P(z) is a preemphasis filter at the transmitter and 1/P (z) is the corresponding deemphasis filter at the receiver. This structure has been commonly used in PCM coding of speech. The main advantage of using preemphasis is the reduction of dynamic range of speech signal to the quantizer. Thus, for a given number of quantizing levels, more accurate quantization would be possible, resulting in higher SQNR. Recent development of the theory of auditory masking shows that preemphasis of speech results in not only improved SQNR but also perceptually pleasing quality because of its noise spectral shaping property [8]. In Fig. 1 the output signal $\hat{S}(z)$ can be shown to be

$$\hat{S}(z) = S(z) + Q(z)/P(z) \qquad (1)$$

where S(z) and Q(z) are the z-domain representations of the input signal and quantization noise, respectively. As a result, the shape of the output noise spectrum depends on 1/P(z) if we assume that Q(z) has flat spectral shape. A fixed preemphasis filter has generally been used whose shape is such that 1/P(z) has the same spectral shape as the long-term average of speech spectrum. This fixed preemphasis filter yields improved performance results over the case without preemphasis (18).

Another method of noise spectral shaping is one that uses a noise feedback filter which can adjust the noise spectrum. A PCM coder with a noise feedback filter is shown in Fig. 2. It can readily be shown that, in the absence of channel errors, the decoded output $\hat{S}(z)$ is given by

$$\hat{S}(z) = S(z) + B(z)Q(z). \qquad (2)$$

Therefore, a simple way to modify the spectrum of quantization noise is to select a filter B(z) that matches the desired spectral shape of noise.

It is important to note that the noise spectral shaping method described here does not require additional information to be transmitted to the receiver even when adaptive noise shaping is done. In the present method noise spectral shaping is done at the transmitter only. Hence, the perceptual improvement can be attained with some increase in the transmitter complexity, but without increase in data rate. Therefore, we will use the latter method of noise shaping to improve the subjective performances of APCM, ADPCM and ADM.

When it is assumed that the quantization noise is white, the spectrum of coder output noise is determined only by the noise feedback filter B(z). The constraint on B(z) to maintain the stability of the feedback loop is given by [6]

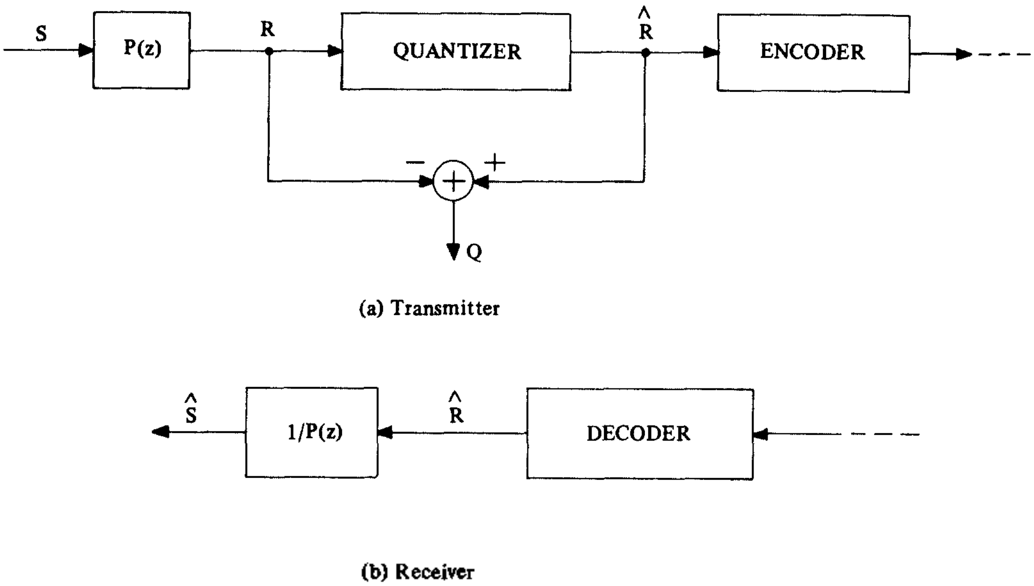


Fig. 1. Block diagram of PCM system with preemphasis and deemphasis

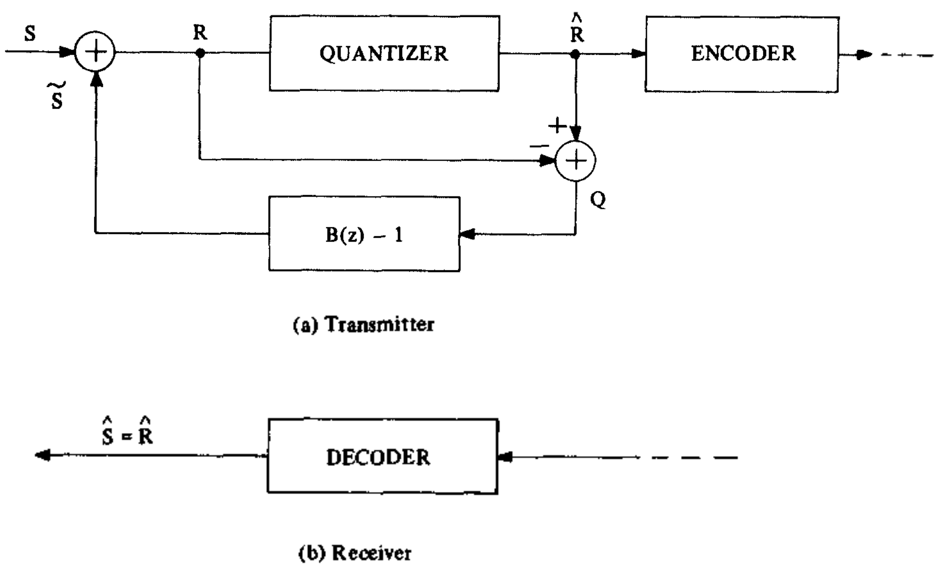


Fig. 2. Block diagram of PCM system with a noise feedback filter

$$\frac{1}{f_s}\int_0^{f_s} \ln |B(f)|^2 \quad df = 0 \qquad (3)$$

where f_s is the sampling frequency, T is the sampling time interval and $B(f)=B(z)|_{z=e^{j2\pi fT}}$.

Assuming that the power of the quantization noise Q(z) is not changed significantly by the feedback loop, the average value of the log power spectrum of output noise is then determined solely by the quantizer and is not altered by the choice of the filter B(z). The filter B(z), however, redistributes the noise power from one frequency band to another. Thus, reduction of quantization noise in one frequency band is achieved at the expense of increasing the quantization noise in another band. Since a large part of perceived noise in a coder comes from the frequency region where the signal level is low, the filter B(z) can be used to reduce noise in the regions of low spectral amplitude, while increasing noise in the regions near formant frequencies where the noise can effectively be masked by speech signal.

Another important condition that the filter B(z) must satisfy is that the expression [B(z)-1] is a predictior that operates only on past values of the noise Q(z), since Q(z) becomes available only after R(z) is computed. Therefore, B(z) should be of the form

$$B(z) = i + \sum_{k=1}^{M} b_k z^{-k} . \qquad (4)$$

Consequently, from (3) and (4), the function B(z) is a minimum phase transfer function with spectrum $|B(f)|^2$.

III. DESIGN OF NOISE FEEDBACK FILTERS

The noise feedback filter B(z) can be designed such that it minimizes an error measure in which the noise is weighted according to some subjectively meaningful criterion. If a subjectively effective noise measure can be defined using a weighting function W(f), then an optimal noise shaping filter B(z) can be chosen. In this case the filter would have the same spectral shape as that of 1/W(f) [6].

To design the noise shaping filter B(z), we can do as follows. First, by Fourier transformation we transform $|1/W(f)|^2$ to obtain an autocorrelation function. Using a procedure similar to LPC analysis, we then compute the autocorrelation function to determine a set of predictor coefficients, thereby getting the desired filter coefficients $\{b_k\}$ for the filter B(z).

Another method to obtain B(z) is approximating the impulse response of 1/W(z) by that of B(z) such that the mean-squared difference of each coefficient is minimized. The coefficients $\{b_k\}$ are then obtained as a solution to a set of linear normal equations by using the linear prediction technique. However, the use of these methods to obtain B(z) has some drawbacks. The first method often results in unstable coefficient sets because the autocorrelation function is not obtained from a real data sequence. Thus, the stability of the feedback loop is not guaranteed The second method seems to be a good candidate to obtain a stable coefficient set for B(z). The method requires the noise weighting function W(z) to be minimum phase. Nevertheless, because the filter order of B(z) is limited to that of W(z), the resulting feedback filter B(z) is generally not adequate to fully approximate 1/W(z) (see Fig. 3.)

In this work, we use a method different from those discussed above to obtain the feedback filter B(z). The resulting B(z) is guaranteed to be stable and the filter order can be adjusted as desired. In addition, the noise weighting function W(z) is not required to be minimum phase. The procedure of designing the shaping filter by this method is as follows.

(1) Determine the magnitude of the frequency response of the desired noise weighting function W(f). The resulting function is a C-message weighting function for APCM and ADPCM or a high-pass filter for ADM.
(2) Design an FIR digital filter that approximates W(f) using the method of McClellan et al [17].
(3) Generate a white Gaussian sequence.
(4) Filter the white Gaussian sequence by the FIR filter designed in Step (2).
(5) Analyze the filtered Gaussian sequence by using the LPC analysis method.
(6) Plot the frequency response of the LPC coefficients. If the result is satisfactory, go to Step (7). Otherwise, increase or decrease the order of LPC analysis and go to Step (5).
(7) The LPC coefficients obtained in Step (5) are the filter coefficients of B(z).

The frequency domain characteristics of the C-message weighting function and the corresponding noise feedback filter are shown in Fig. 3. In this case, the z-domain transfer function of the C-message weighting function is given by [19]

$$W(z) = 1 - 0.59z^{-1} - 0.39z^{-2} \qquad (5)$$

Hence, instead of designing an FIR filter which approximates W(f) from the procedure of obtaining B(z) discussed above, we used (5) directly to obtain the filtered Guassian sequence.

As for ADM, the desired characteristics of the feedback filter is such that the filter shifts some amount of quantization noise to a region outside the baseband. Accordingly, the noise feedback filter B(z) in ADM is quite different from those used in APCM and ADPCM. The desired characteristics of the noise shaping filter for ADM is shown in Fig. 4. Note that the benefit actually achievable by using this type of a feedback filter is limited because in ADM noise shifting causes an increase of slope-overload distortion. According to our simulation results, the feedback filter that has about 6 dB difference in magnitude between in-band and out-band (see Fig. 4) resulted in nearly optimal performance. Following the design procedure for a noise feedback filter B(z) discussed above, we have obtained noise shaping filters for ADM whose frequency domain characteristics are shown in Fig. 5 for the clock rates of 16, 24 and 32 kHz. The number of coefficients of those noise feedback filters in all cases was 14. The values of the coefficients of the noise feedback filters designed for APCM, ADPCM and ADM are given in Table 1.

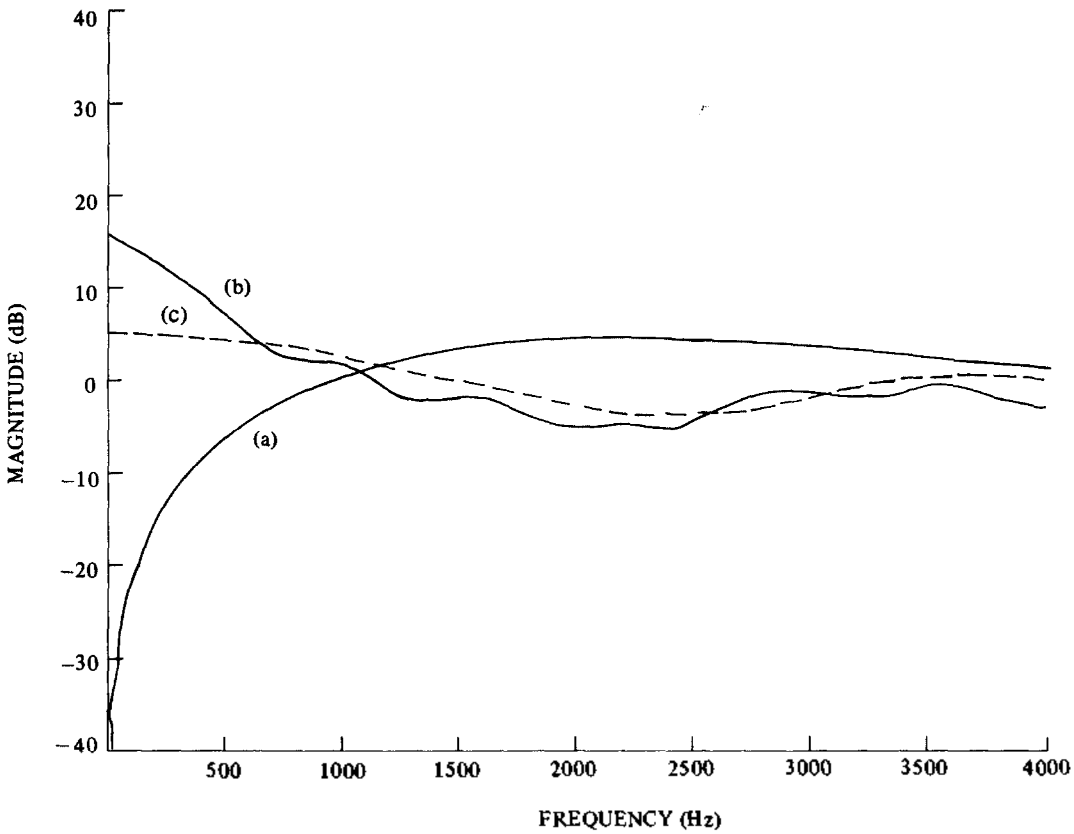


Fig. 3. (a) Frequency response of noise weighting function (C-message weighting), (b) frequency response of the corresponding noise feedback filters obtained from filtered Gaussian sequence (solid line) and (c) frequency response obtained by the method of approximating the impulse response of 1/W(z)

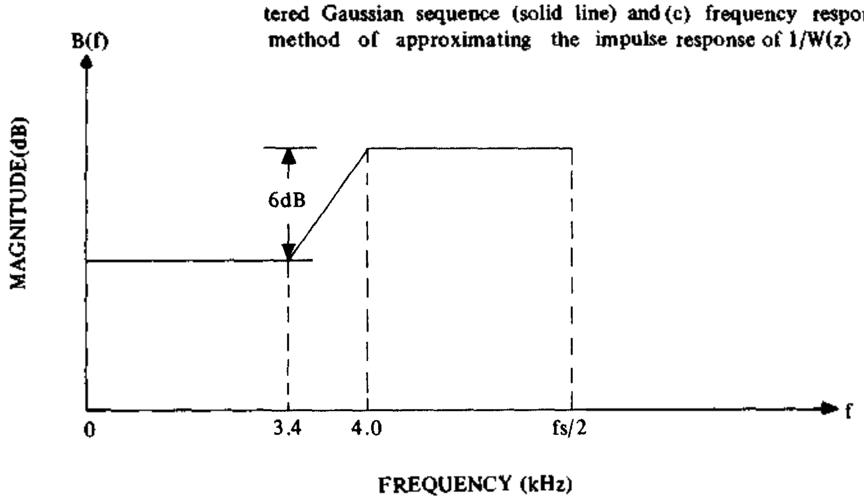
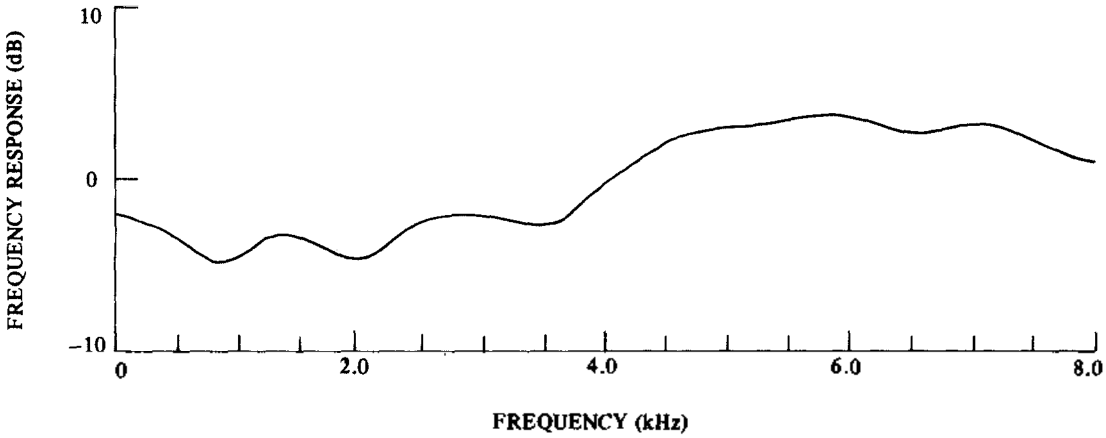
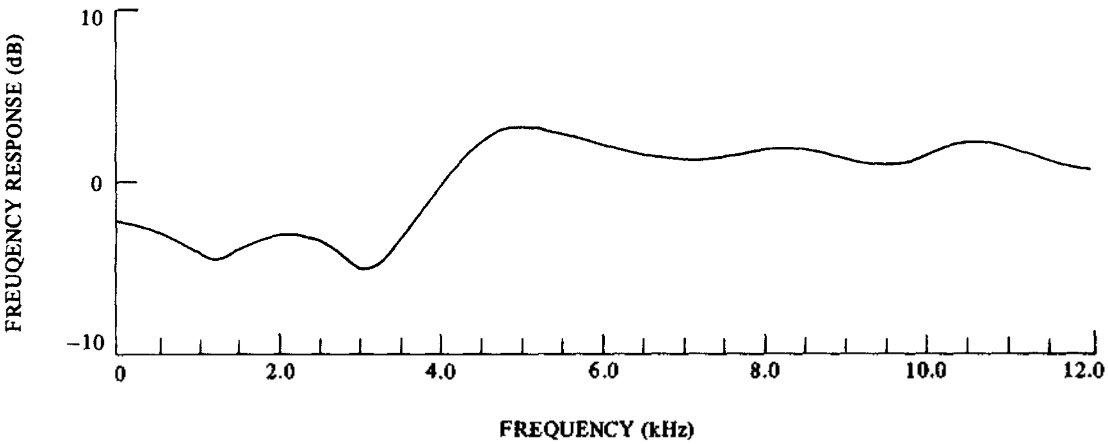


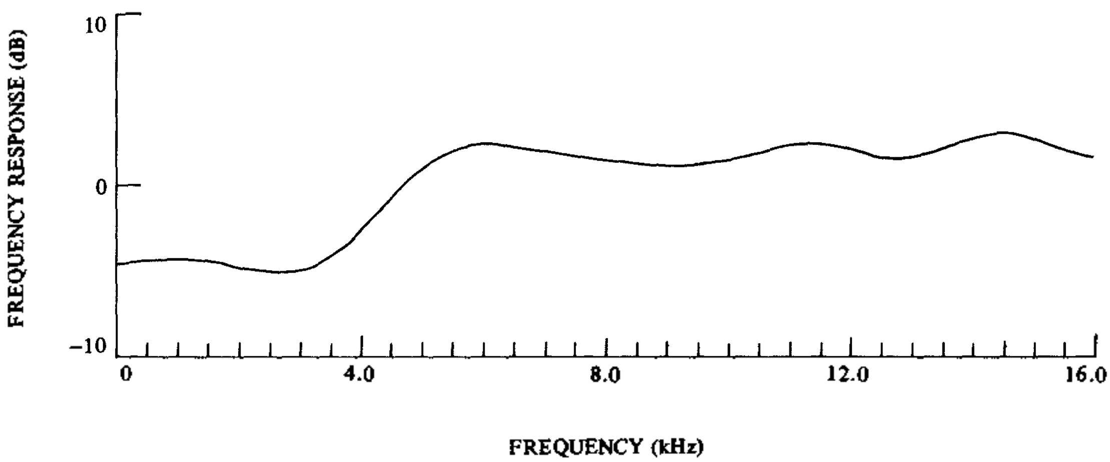
Fig. 4. Frequency domain characteristics of the desired noise feedback filter B(z) for ADM



(a)



(b)

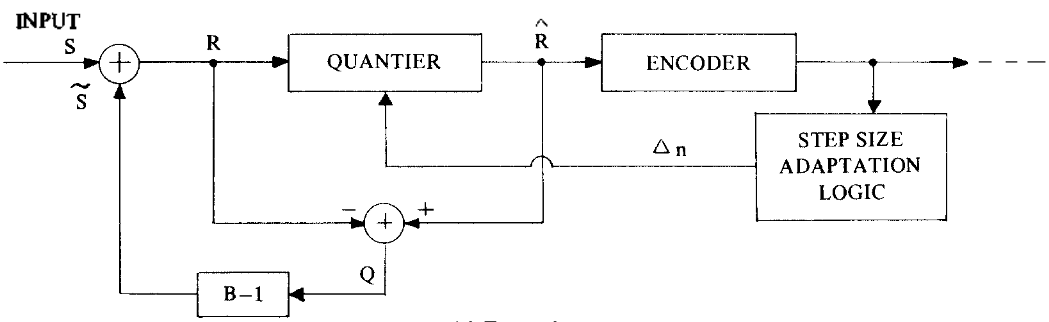


(c)

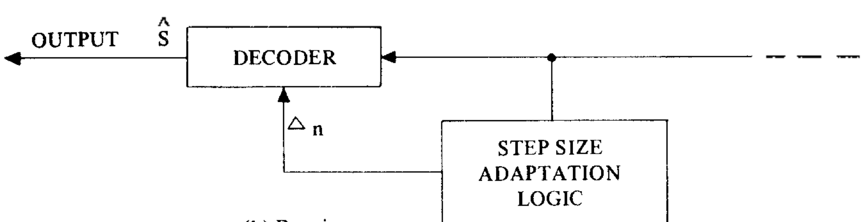
Fig. 5 Frequency response of the noise feedback filter B(z): (a) f_s = 15kHz (b) f_s = 24kHz (c) f_s = 32 kHz

Table 1. Noise feedback filter coefficients

CODER TYPE / COEFFICIENT	APCM, ADPCM (f_s = 8 kHz)	ADM (f_s = 16kHz)	ADM (f_s = 24 kHz)	ADM (f_s = 32 kHz)
b_1	.5538	−.4110	−.3382	−.2698
b_2	.6745	4.228×10^{-2}	−.1364	−.1745
b_3	.5496	.1396	3.712×10^{-2}	-6.515×10^{-2}
b_4	.5243	-1.714×10^{-2}	8.879×10^{-2}	2.026×10^{-2}
b_5	.4833	-6.246×10^{-2}	7.356×10^{-2}	8.065×10^{-2}
b_6	.4477	1.555×10^{-2}	5.557×10^{-3}	7.593×10^{-2}
b_7	.3837	3.760×10^{-2}	-6.107×10^{-2}	1.125×10^{-2}
b_8	.3138	-5.441×10^{-2}	-4.061×10^{-2}	-3.753×10^{-2}
b_9	.3122	1.777×10^{-2}	6.398×10^{-2}	7.290×10^{-3}
b_{10}	.2444	-2.865×10^{-2}	2.410×10^{-2}	-5.635×10^{-2}
b_{11}	.2290	7.012×10^{-2}	6.578×10^{-2}	4.911×10^{-3}
b_{12}	.1698	5.325×10^{-3}	1.863×10^{-3}	5.344×10^{-5}
b_{13}	.1087	-1.860×10^{-2}	-2.379×10^{-2}	8.847×10^{-3}
b_{14}	3.967×10^{-2}	-3.001×10^{-2}	-1.239×10^{-2}	7.819×10^{-4}



(a) Transmitter



(b) Receiver

Fig. 6 Block diagram of APCM-NS

IV. WAVEFORM CODERS WITH NOISE SPECTRAL SHAPING

A. Adaptive PCM with Noise Spectral Shaping (APCM-NS)

The structure of APCM-NS that combines PCM-NS (see Fig. 2) and a step size adaptation logic is shown in Fig. 6. When no channel error is present, the output signal of the APCM-NS system is given by (2).

The step size adaptation logic for an adaptive quantizer of APCM-NS is the same as that of a conventional APCM. Its operation can be summarized as follows. In Fig.6, input speech samples s_n are quantized and encoded for binary transmission. Also, based on the output bit stream $\{b_n\}$, the step size adaptation logic generates a basic step size Δ_n at time instant n to adjust its quantizer step size. There are a number of strategies for step size adaptation. Of those, as mentioned earlier, two types of adaptation schemes are used here. One scheme is the instantaneous companding method. In this case the quantizer basic step size changes instantaneously at each sampling time in response to the bit pattern just transmitted. The step size adaptation strategy used is based on the one-word memory approach proposed by Jayant [20]. The algorithm is as follows. The coder input signal s_n is quantized to one of 2^b levels. The step size adaptation logic examines the quantizer output bits for the n-th sample and computes the quantizer step size, Δ_{n+1}, for the (n+1)-th sample according to the relation

$$\Delta_{n+1} = \Delta_n M(|H_n|) \qquad (6)$$

and

$$\Delta_{min} \leq \Delta_{n+1} \leq \Delta_{max}, \qquad (7)$$

where Δ_n is the step size used for the n-th sample, and $M(|H_n|)$ is a multiplication factor whose value depends on the quantizer magnitude level H_n at time n. It can take on one of 2^{b-1} values; M_1, M_2, . . . , and $M_{2^{b-1}}$. Typical values of M_i for 4-bit APCM coders are given in [20].

The other step size adaptation scheme is the syllabic companding method. The adaptation algorithm used in syllabically companding APCM-NS coder adjusts the magnitude of the basic step size Δ_n of the quantizer at a much slower rate compared with instantaneous variations of input speech signal. The step size adaptation logic adjusts its quantizer step size based on the output bit patterns $\{b_n\}$. When the mean value of L consecutive quantizer mangitude levels $\{|H_n|\}$ exceeds the given threshold level, the logic applies the voltage Δ_{max} to an RC integrator with a time constant $1/\alpha$ s. Then, the step size (i.e., the output of the RC integrator) increases towards the maximum step size Δ_{max}. On the other hand, for the input of lower amplitudes, the above condition cannot be met. When the adaptation logic does not acquire the condition of step size increase, the logic applies no voltage to the RC integrator, thus decreasing the step size Δ_n towards Δ_{min}. Step size adaptation is controlled by a time constant $1/\alpha$ which is of the order of several milliseconds. The adaptation rule can be written as follows:

$$\Delta_n = e^{-\alpha\tau}\Delta_{n-1} + \Delta_{max}(1-e^{-\alpha\tau}),$$

$$\text{if } \frac{\sum_{i=1}^{L} |H_{n-i}|}{2^{b-1}\, L} > C$$

$$= e^{-\sigma T}\Delta_{n-1}, \text{ otherwise} \qquad (8)$$

where T is sampling interval and Δ_n is larger than or equal to the minimum step size Δ_{min}. Note that the step size adaptation scheme of the syllabically companding APCM is similar to that of the continuously variable slope DM (CVSD) [22]. The only difference between them is how to detect the threshold point (i.e., C in (8)) at which the step size begins to increase. The value of L should be chosen according to channel conditions such that L should be larger as channel errors increase.

Because speech signal is not uniformly distributed, a nonuniform quantizer that approximates distributions of input speech signal should be used to improve further the APCM performance. Generally, the distribution of real speech signal has been assumed to be Gaussian, gamma, or Laplacian. In this study, we use the gamma quantizer in APCM and ADPCM becasue it yields the best performance.

B. Adaptive DPCM with Noise Spectral Shaping (ADPCM-NS)

The structure of an ADPCM-NS system is shown in Fig. 7. In the figure it can be easily shown that the output speech is given by (2) and the input to the quantizer is

$$R(z) = S(z) - \hat{R}(z)P(z) / [1-P(z)] + [B(z)-1]Q(z) \qquad (9)$$

where P(z) is a prediction filter and R(z) is the quantizer output.

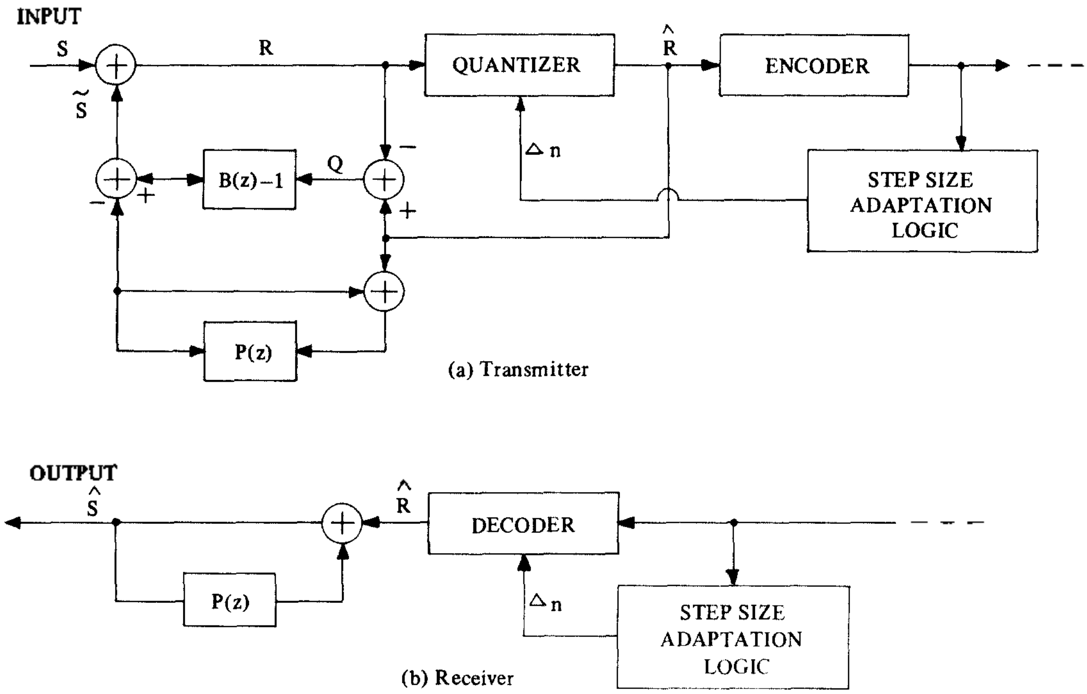
The step size adaptation scheme of the ADPCM-NS system is basically the same as that of the APCM-NS system. Without the noise feedback loop the ADPCM-NS system becomes the same as the ADPCM system originally proposed by Cummiskey et al.[22] The only difference is the step size multipliers that are tailored to match the difference signal of speech when we use an ADPCM-NS with instantaneous companding [20]. For the prediction filter in the feedback path of the ADPCM-NS coder, we use a first-order fixed predictor.

C. Adaptive DM with Noise Spectral Shaping (ADM-NS)

The basic structure of ADM-NS can be obtained from the structure of ADPCM-NS as shown in Fig. 8. It includes a one-bit quantizer, a first-order fixed predictor P(z), a noise shapig filter B(z), and a decoder. The decoder derives the current step size, and hence the quantized signal R(z) by an adaptation algorithm specific to the ADM algorithm chosen. In this study, we use the CVSD and the first-order constant factor DM (CFDM) for the noise spectral shaping study of ADM. A detailed description of the CVSD and the CFDM systems is given in [21].

V. SIMULATION RESULTS AND DISCUSSION

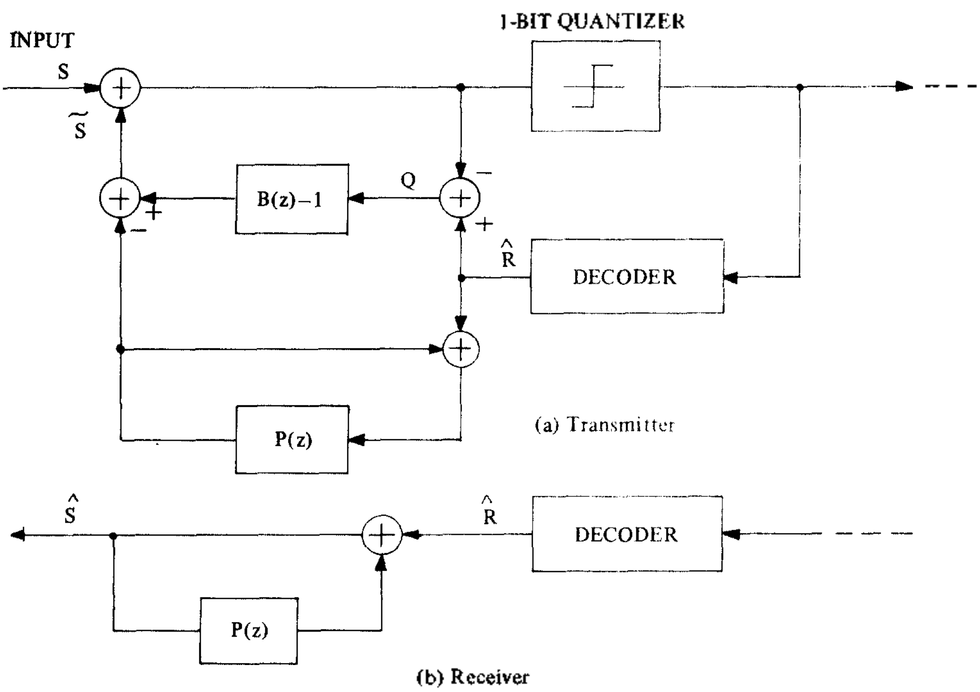
Computer simulations of the performances of APCM-NS, ADPCM-NS, and



(a) Transmitter

(b) Receiver

Fig. 7 Block diagram of ADPCM-NS



(a) Transmitter

(b) Receiver

Fig. 8 Block diagram of ADM-NS

ADM-NS have been performed. Speech samples used for simulations were made using five sentences spoken by three male and two female speakers and band-limited to 3.4 kHz. The total length of the speech samples was about 15 s. The noise shaping coders under study were simulated with a step size range of 60 dB, and the input signal level was varied over a range of 80dB. The time constants of the syllabic filter and the first-order fixed predictor were set to 4 ms and 1 ms, respectively.

The method of obtaining FWSQNR for APCM-NS and ADPCM-NS is shown in Fig.9. The FSWQNR is defined by

$$\mathrm{FWSQNR}=\frac{\sum_{n=1}^{n=M} s_n^2}{\sum_{n=1}^{n=M} \hat{e}_n^2} \tag{10}$$

Where $\{s_n\}$ is the input speech sample sequence, $\{e_n\}$ is the frequency-weighted error sequence, and N is the number of frames of input speech and M is the number of speech samples in one frame. The weighted error sequence $\{e_n\}$ can be obtained by filtering the difference between the input sequence $\{s_n\}$ and the reproduction $\{\hat{s}_n\}$ using a frequency weighting filter W(z) and a band-limiting low-pass filter. Also, we have used the frequency-weighted segmental SQNR (FWSQNR_{SEG}) as

$$\mathrm{FWSQNR}_{SEG}(\mathrm{dB})=\frac{10}{N}\sum_{i=1}^{N}\log_{10}(1+\mathrm{FWSQNR}_i) \tag{11}$$

where FWSQNR_i is the FWSQNR of the i-th frame of speech. In this study, we employ a C-message weighting function for frequency weighting whose z-domain representation is given approximately by (5). The frequency response of the weighting function is given by

$$|W(f)|^2=1.5002-0.7198\cos 2\pi f+0.78\cos 4\pi f \tag{12}$$

and is plotted in Fig. 3. Note that the high frequency region of the error spectrum is emphasized, whereas the region below about 1 kHz is deemphasized.

The performances in FWSQNR and FWSQNR_{SEG} of APCM-NS and APCM are shown in Figs.10 and 11 respectively. As seen in the figures, APCM-NS with instantaneous companding yields 1 to 1.5 dB performance improvement both in FWSQNR and FWSQNR_{SEG} over the conventional APCM. As for APCM-NS with syllabic companding, the performance gain is about 1.5 to 2 dB over APCM. This gain becomes larger for lower input signal level.

When noise spectral shaping is incorporated in ADPCM, a performance gain in FWSQNR and FWSQNR_{SEG} of about 1 to 1.5dB can be achieved. This is shown in Figs. 12 and 13. To show the effectiveness of noise spectral shaping, the spectra of the original speech and the quantization noise with and without noise spectral shaping for APCM and ADPCM are plotted in Figs. 14 and 15, respectively. The effectiveness of noise spectral shaping can be seen in the figures. Clearly, the noise in the frequency region of 1 to 3 kHz has been reduced.

The noise shaping filter used in ADM is different from that of APCM-NS or ADPCM-NS. Becasue it shifts some of the in-band noise out of the baseband, SQNR is improved.

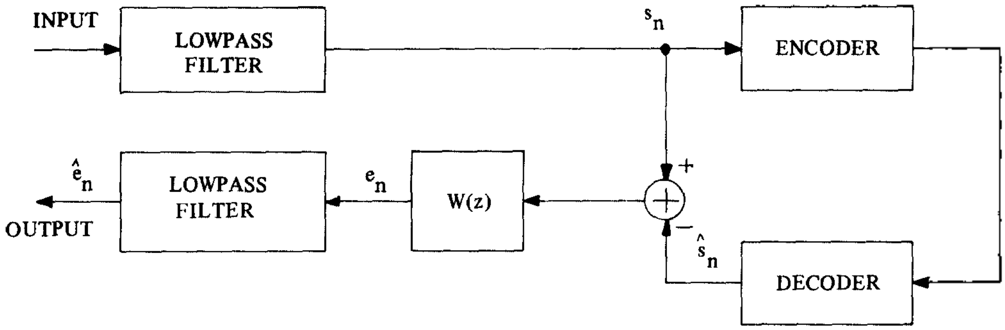


Fig. 9 Block diagram of the procedure for computing frequency-weighted quantization noise for APCM-NS and ADPCM-NS

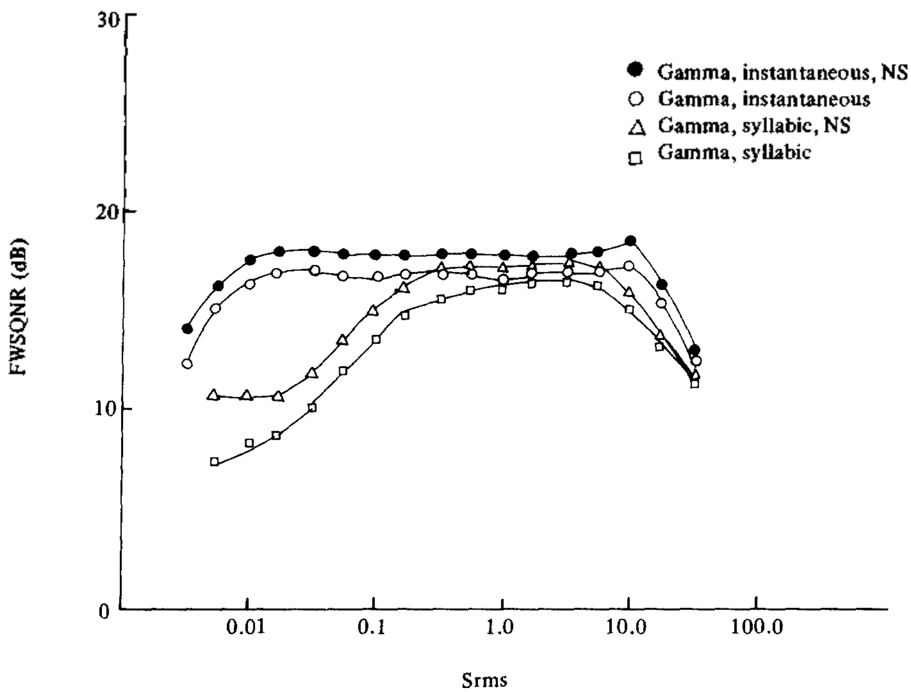


Fig. 10 FWSQNR of APCM systems vs. input signal level (C-message weighting, b = 4)

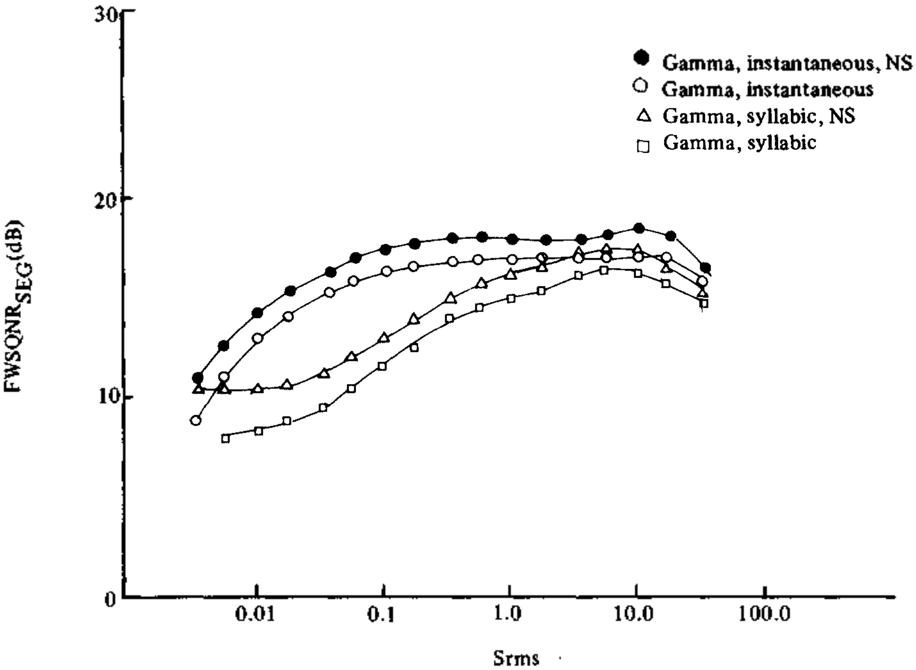


Fig. 11 FWSQNR$_{SEG}$ of APCM systems vs. input signal level (C-message weighting, b=4)

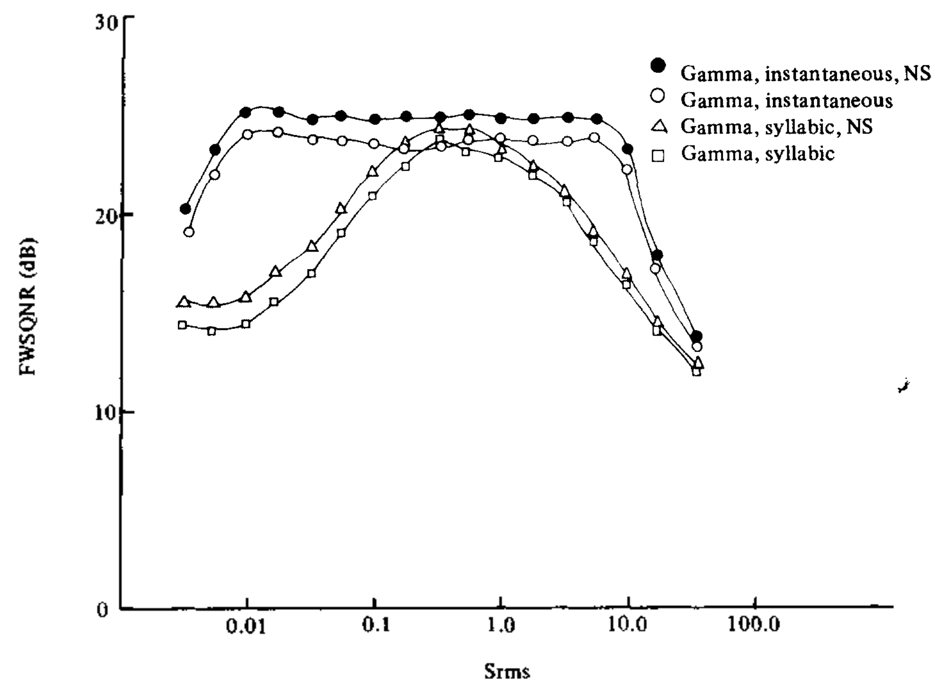


Fig. 12 FWSQNR of ADPCM systems vs. input signal level (C-message weighting, b = 4)

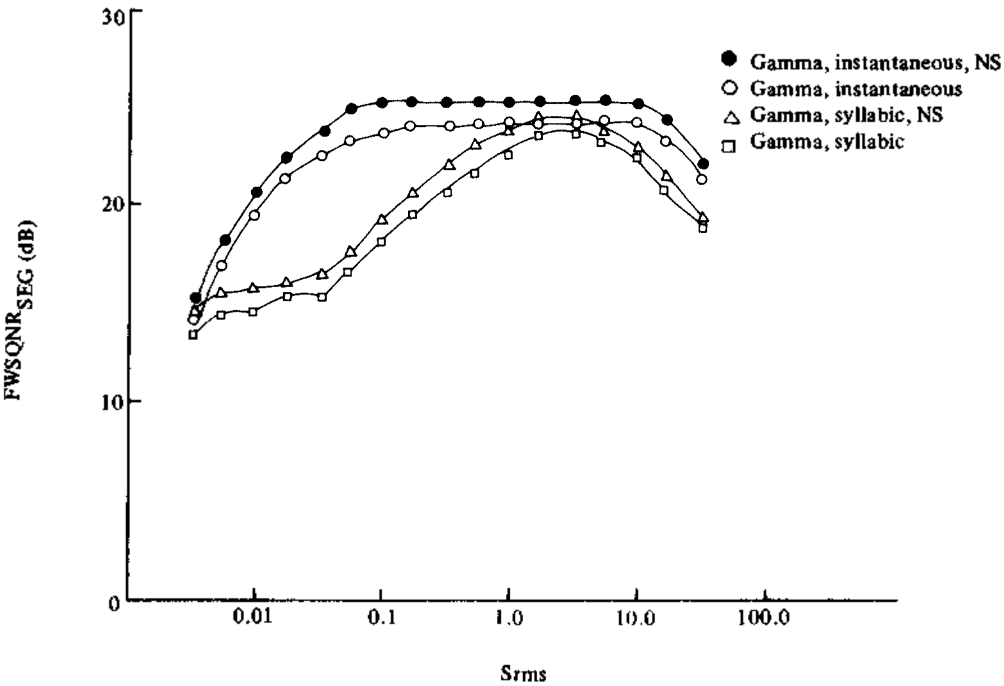


Fig. 13 FWSQNR$_{SEG}$ of ADPCM systems vs. input signal level (C-message weighting, b = 4)

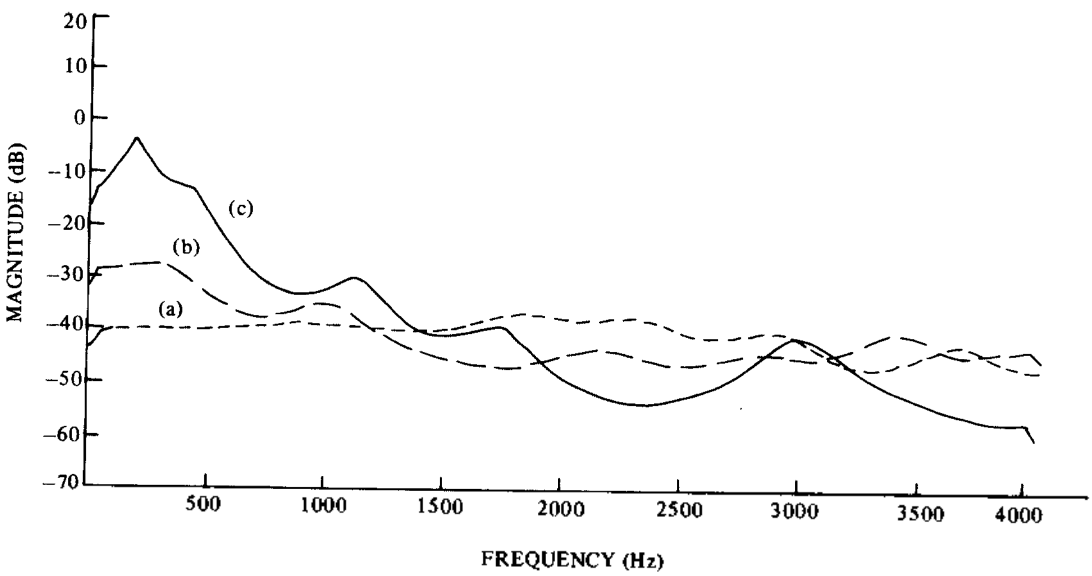


Fig. 14 Comparison of spectral envelopes of (a) output quantization noise in APCM, (b) output quantization noise in APCM-NS and (c) the corresponding input speech (b=4)

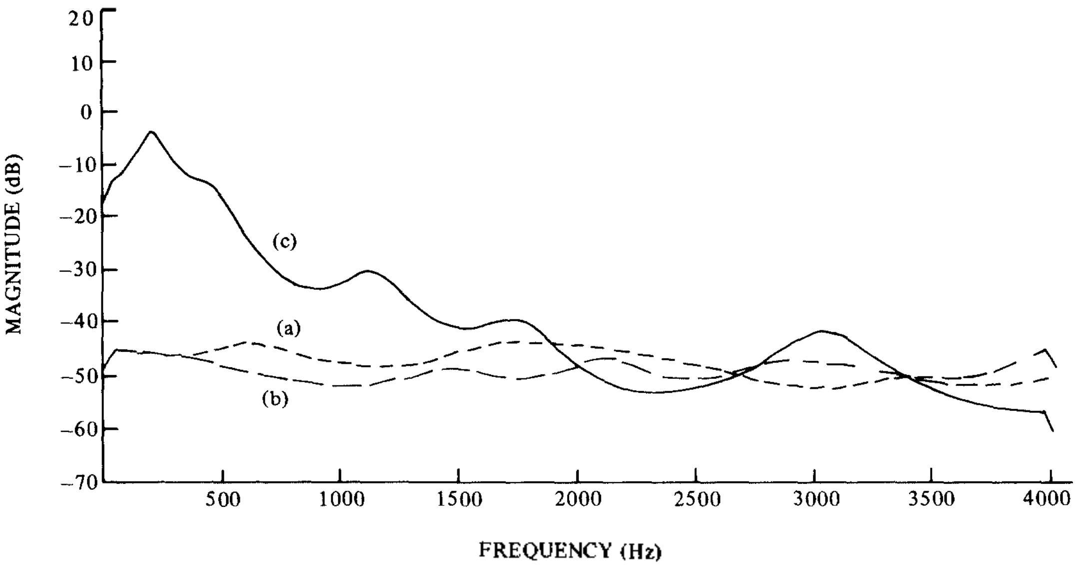


Fig. 15 Comparison of spectral envelope magnitudes of (a) output quantization noise in ADPCM, (b) output quantization noise in ADPCM-NS and (c) the corresponding input speech (b = 4)

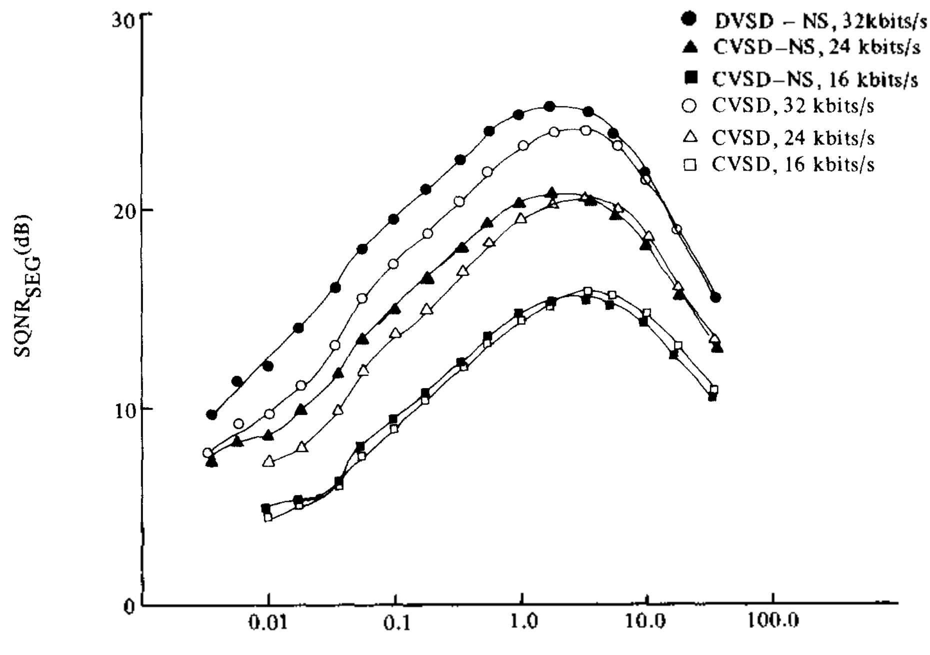


Fig. 16 $SQNR_{SEG}$ of CVSD-NS and CVSD vs. input signal level (f_s = 16, 24 and 32 kHz, f_c = 3.4 kHz)

The performance in $SQNR_{SEG}$ of CVSD-NS is shown in Fig. 16. The clock rates are 16, 24 and 32 kHz. Compared with the performance of the conventional CVSD, a $SQNR_{SEG}$ gain of about 0.5dB at the clock rate of 16 kHz, 0.5 to 2dB at 24 kHz, and 1 to 3 dB at 32 kHz could be obtained.

On the other hand, as shown in Fig. 17, the performance of CFDM-NS at the bit rates of 16 and 24 kbits/s slightly degrades compared with the conventional CFDM due to the increase of slope overload noise caused by noise shaping. An improvement of 0.5 to 2.5dB in $SQNR_{SEG}$ could be obtained for CFDM-NS only at the clock rate of 32 kHz.

Like APCM-NS, the performance gain of noise shaping in ADM-NS also becomes larger when the input signal level is lower. The spectra of the original speech and the quantization noise in CVSD and CVSD-NS at the bit rate of 16 kbits/s are shown in Fig. 18. The effect of the noise shaping in reducing the baseband noise can be seen.

From the performance results discussed above, one can conclude that noise spectral shaping is particularly effective when it is incorporated into a waveform coder with syllabic companding. It is noted that the quantization noise of a waveform coder with syllabic companding is largely granular noise when the signal level is low, and that the basic assumption in designing a noise feedback filter is that the coder is in granular noise region. Also note that granular noise is subjectively more annoying than overload noise [23], [24]. Therefore, the use of a noise feedback filter for noise spectral shaping in coding speech should result in perceptually pleasing sound. According to our informal listening tests, although the performance improvement in objective measures was relatively small, the waveform coders with noise spectral shaping did result in perceptively more pleasing and intelligible sound than the coders without it. The differences between them were perceptible.

VI. CONCLUSIONS

A detailed study on the use of noise spectral shaping in APCM, ADPCM, and ADM systems has been done. Two kinds of noise spectral shaping have been considered. One has been devised based on auditory masking of quantization noise. In this scheme the in-band quantization noise is spectrally shaped according to a certain frequency-weighted error criterion. This method has been applied to spectral noise shping of APCM and ADPCM systems. The frequency-weighted error criterion has been chosen to minimize the C-message weghted noise. For ADM systems whose sampling rate is in general several times the Nyquist rate, another method of noise spectral shaping has been used to shift some portion of in-band noise outside the signal band. In either case, a new method has been proposed to obtain the noise feedback filters such that the filter coefficients are obtained from the filtered Gaussian sequence using the autocorrelation linear prediction analysis method. The resulting noise filters are guaranteed to be stable.

Computer simulations of APCM, ADPCM, and ADM with noise spectral shaping show that the improvement of the performance

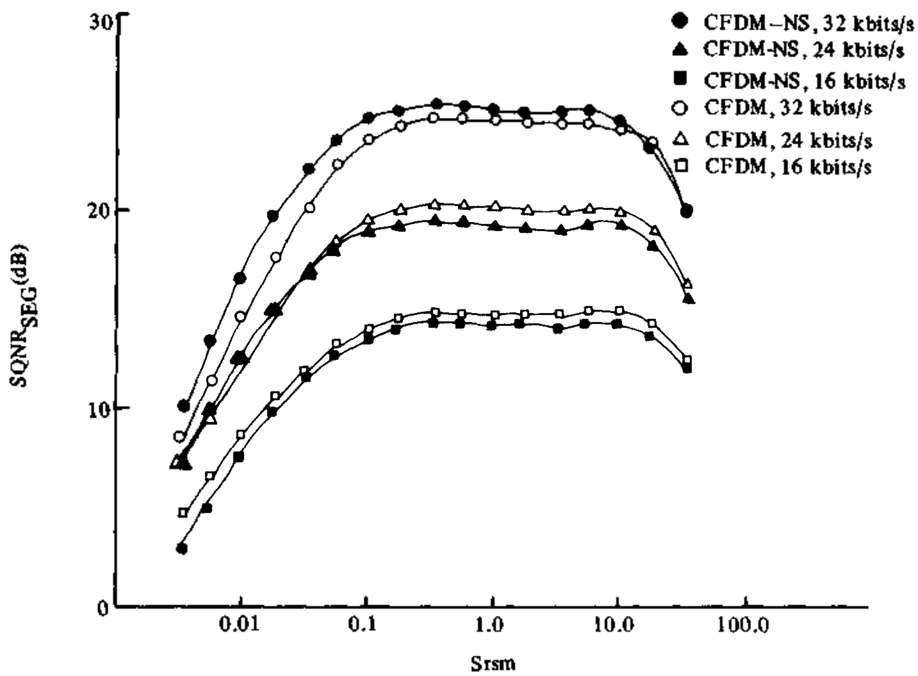


Fig. 17 $SQNR_{SEG}$ of CFDM-NS and CFDM vs. input signal level (f_s = 16, 24 and 32 kHz, f_c = 3.4 kHz)

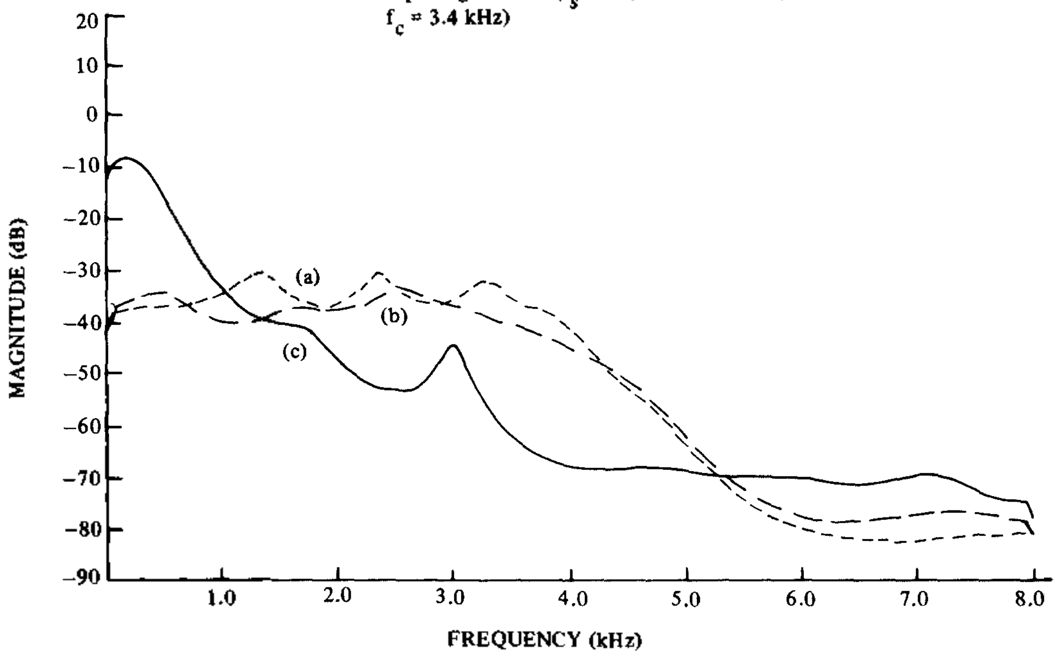


Fig. 18 Comparison of spectral envelope magniudes of (a) output qnantization noise in CVSD, (b) output quantization noise in CVSD-NS and (c) the corresponding input speech (f_s = 16 kHz, f_c = 3.4 kHz)

of each coder is about 0.5 to 3 dB over the coders without it. Although this improvement is relatively small, those coders with noise spectral shaping yielded perceptually pleasing sound that can be readily distinguished from that without shaping. The improvement is greater when the noise spectral shaping is applied to waveform coders with syllabic companding than those with instantaneous companding.

REFERENCES

(1) M.R. Schroeder, B.S. Atal and J.L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," J. Acoust. Soc. Amer., vol. 66, pp. 1647-1652, Dec. 1979.

(2) B. J. McDermott and C. Scagliola, "The perception of spectrally shaped additive noise in speech," in Proc. 1982 IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 196-198, 1982.

(3) P. Noll, "On predictive quantizing schemes," Bell Syst. Tech. J., vol. 57, pp. 1499-1532, May-June 1978.

(4) H.A. Spang, III and P.M. Schultheiss, "Reduction of quantizing noise by use of feedback," IRE Trans. Commun. Syst., vol. CS-10, pp. 373-380, Dec. 1962.

(5) E.G. Kimme and F.F. Kuo, "Synthesis of Optimal filters for a feedback quantization system," IEEE Trans. Circuit Theory, vol. CT-10, pp. 405-413, Sept. 1963.

(6) B.S. Atal and M.R. Schroeder, "Predictive coding of speech signals and subjective error criteria," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, pp. 247-254, June 1979.

(7) B.S. Atal, "Predictive coding of speech at low bit rates," IEEE Trans. Commun., vol. COM-30, pp. 600-614, Apr. 1982.

(8) J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding in predictive coding of speech," IEEE Trans. Acoust., Speech, Signal processing, vol. ASSP-27, pp. 63-73, Feb. 1979.

(9) J. Makhoul and M. Berouti, "Predictive and residual encoding of speech," J. Acoust. Soc. Amer., vol. 66, pp. 1633-1641, Dec. 1979.

(10) R. Bastian, "Subjective improvements in DPCM-AQ performance based on adaptive noise shaping," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-29, pp. 1067-1071, Oct. 1981.

(11) M. Copperi, et al., "An adaptive delta modulator with noise spectral shaping," in Proc. 1981 IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 828-831, 1981.

(12) N.R. French and J.C. Steinburg, "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Amer., vol. 19, pp. 90-119, Jan. 1947.

(13) S.S. Stevens, "Perceived level of noise by Mark VII and decibels (E)," J. Acoust. Soc. Amer., vol. 51, pp. 575-600, Feb. 1971.

(14) W.T. Cochran and D.A. Lewinski, "A new measuring set for message circuit noise," Bell Syst. Tech. J.,

vol. 39, pp. 911-931, July 1960.

(15) A.J. Aikens and D.A. Lewinski, "Evaluation of message circuit noise," Bell Syst. Tech. J., vol. 39, 879-910, July 1960.

(16) G.A. Miller, "The masking of speech," Psychol. Bull., vol. 44, pp. 105-129, Mar. 1947.

(17) J.H. McClellan, T.W. Parks and L.R. Rabiner, "A computer program for designing optimum FIR linear phase digital filters," IEEE Trans. Audio, Electroacoust., vol. AU-21, pp. 506-526, Dec. 1973.

(18) J.D. Gibson, "Aadaptive prediction in speech differential encoding systems," Proc. IEEE, vol. 68, pp. 488-527, Apr. 1980.

(19) S.G. Wilson and S. Husain, "Adaptive tree encoding of speech at 8000 bits/s with a frequency weighted error criterion," IEEE Trans. Commun., vol. COM-27, pp. 165-170, Jan. 1979.

(20) N.S. Jayant, "Adaptive quantization with a one-word memory," Bell Syst. Tech. J., vol. 52, pp. 1119-1144, Sept. 1973.

(21) C.K. Un and H.S. Lee, "A study of the comparative performance of adaptive delta modulation systems," IEEE Trans. Commun., vol. COM-28, pp. 96-101, Jan. 1980.

(22) P. Cummiskey, N.S. Jayant and J.L. Franagen, "Adaptive quantization in differential PCM coding of speech," Bell Syst. Tech. J., vol. 52, pp. 1105-1118, Sept. 1973.

(23) H. Levitt, C.A. McGonegal and L.L. Cherry, "Perception of slope-overload istortion in delta-modulated speech signals," IEEE Trans. Audio Electroacoust., vol. AU-18, pp. 240-247, Sept. 1970.

(24) N.S. Jayant and A.E. Rosenberg, "The preference of slope overload to granularity in the delta modulation of speech," Bell Syst. Tech. J., vol. 50, pp. 3117-3125, Dec. 1971.