PRE-FILTER와 POST-FILTER를 사용하는 음성파형 부호화 방법에 관하여

On the Use of Pre- and Post-Filters in Speech Waveform Coding

*조 동 호 (Cho D.H.)
**은 종 관 (Un C.K.)
***김 제 우 (Kim J.W.)

요 약

이 논문에서는 frequency-weighted MSE를 최소화하는 적응 pre-filter와 post-filter를 음성파형 부호화기에 적용했을 때의 성능을 분석한다. 먼저 여러 다양한 pre-filter와 post-filter에 의한 noise shaping 효과를 이론적으로 보여준다. 그리고 frequency-weighted SNR척도를 사용하여 적응 pre-filter와 post filter에 의한 성능면에서의 이득을 이론적으로 유도한다. 적응 pre-filter와 post- filter를 ADM과 ADPCM 부호화기에 적용해본 결과에 의하면 음성파형 부호하기의 성능을 $FWSNR_{SEG}$ 척도로 약 3dB 정도 개선할 수 있음을 알 수 있다. 또한 pre-filter와 post-filter를 사용하면 청각적으로 중요한 영향을 미치는 1 kHz에서 3 kHz 사이의 양자화 잡음을 효율적으로 줄일 수 있다.

ABSTRACT

In this paper, the performance of adaptive pre- and post-filters that minimize the frequency-weighted mean-squared error is analyzed when they are used in a speech waveform coder. The noise shaping effect by various pre- and post-filters is shown. Then, the frequency-weighted signal-to-quantization noise ratio (FWSNR) gain due to adaptive pre- and post-filtering is derived analytically. Also, computer simulation results for adaptive differential pulse code modulation (ADPCM) and adaptive delta modulation (ADM) are presented. According to the results, the performance can be improved by about 3 dB in segmented frequency-weighted signal-to-noise ratio ($FWSNR_{SEG}$). In addition, it is shown that the quantization noise in the perceptually important frequency region of 1 to 3 kHz can be reduced by using the pre- and post-filters.

I. INTRODUCTION

It has long been recognized that the perceptual effect of noise varies 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]. However, most of waveform coders have been designed to minimize the unweighted mean-squared error between the system input and output. In other words, conventional waveform coders have not been designed to minimize

* 한국과학기술원 통신공학연구실 선임연구원
** 한국과학기술원 전기 및 전자공학과 교수
*** 한국과학기술원 전기 및 전자공학과 박사과정

the perceptual effects of noise. Recent experimental results show that the subjective loudness of quantization noise can be reduced by noise spectral shaping [2], [3].

The spectral shaping of quantization noise may be divided into two methods. One method is the pre- and post-filtering approach to improve the subjective quality of speech. This method was used to improve the quality of pulse code modulation (PCM) and differential PCM (DPCM) [4]. According to the previous results, significant improvement in subjective performance could be achieved for PCM when an optimum mean-square filter was used. However, no significant improvement could be made for DPCM. Also, the preemphasis and deemphasis scheme that may be regarded as a simplified version of pre- and post-filtering has been used successfully in linear predictive coding (LPC) [5]. The other method is the noise feedback coding (NFC) scheme proposed by Noll [6]. This scheme is based on the idea of Cutler [7]. The use of noise feedback circuit was first attempted to reduce the quantization noise of speech coder [8]. So far, noise spectral shaping has been done to improve the subjective qualities of adaptive predictive coder (APC) [9] and adaptive DPCM (ADPCM) [10]. For adpative delta modulation (ADM) systems, noise shaping has been applied to reduce the quantization noise power in the baseband [11].

In this study, a simple pre- and post-filtering scheme that minimizes the mean-squared error or frequency-weighted mean-squared error (FWMSE) is applied to waveform coders such as ADPCM and ADM to improve the subjective quality of speech, and the gain resulting from the use of pre- and post-filtering is analyzed. For a frequency-weighted objective measure, we use a C-message weighting function, which emphasizes high frequency noise relative to low frequency noise [12].

Following this introduction, noise shaping is shown and discussed for different versions of the pre- and post-filters in Section II. In Section III, the performance gain due to adaptive pre- and post-filtering to minimize frequency-weighted MSE is analyzed for real speech. In Section IV, computer simulation results for ADPCM and ADM are presented. Finally, conclusions are made in Section V.

II. SPECTRAL SHAPING OF QUANTIZATION NOISE

The basic structure of an ADPCM system with pre- and post-filters (ADPCM-PPF) is shown in Fig. 1. Without the pre- and post-filters the system is the same as the ADPCM proposed by Cummiskey et al. [13]. When there exist no channel errors, the z-transform of the receiver output signal $\hat{s}(t)$ of the ADPCM-PPF is given by

$$\hat{S}(z) = \frac{\hat{R}(z)}{1 - F(z)} D(z), \quad (1)$$

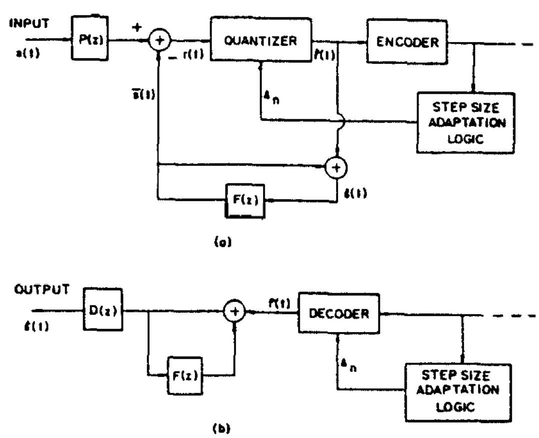

Fig. 1. Block diagram of ADPCM-PPF.
(a) Transmitter (b) Receiver

where F(z) and D(z) are the z-domain representations of the prediction and post-filters, respectively. Also, the z-transform of the quantizer output $\hat{r}(t)$ is given by

$$\hat{R}(z) = R(z) + Q(z), \tag{2}$$

where R(z) and Q(z) are the z-transforms of the quantizer input and the quantization noise, respectively. Hence, the input to the quantizer is expressed as

$$R(z) = S(z)P(z) - \frac{\hat{R}(z)}{1 - F(z)}F(z), \tag{3}$$

where P(z) and S(z) are the z-domain representations of the pre-filter and the input signal, respectively. Combining (1), (2) and (3), we have the output signal

$$\hat{S}(z) = S(z) + D(z)Q(z). \tag{4}$$

Consequently, we can see from (4) that quantization noise Q(z) can be shaped by the pre- and post-filters.

To obtain a reduced structure of the ADPCM-PPF, we note that the output speech is given by (4). At the receiver, the output signal for the ADPCM-PPF system can be shown to be

$$\hat{S}(z) = \frac{\hat{R}(z)}{1 - F(z)}. \tag{5}$$

Using (2), (4) and (5), we obtain the input to the quantizer as

$$R(z) = S(z) - \hat{S}(z)F(z) + \{D(z) - 1\}Q(z). \tag{6}$$

The structure of the resulting ADPCM-PPF is shown in Fig. 2.

Another modified scheme of the ADPCM-PPF is shown in Fig. 3. If there are no channel

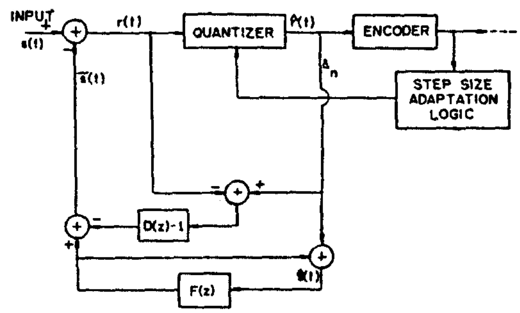

Fig. 2. Block diagram of a simplified ADPCM-PPF transmitter.

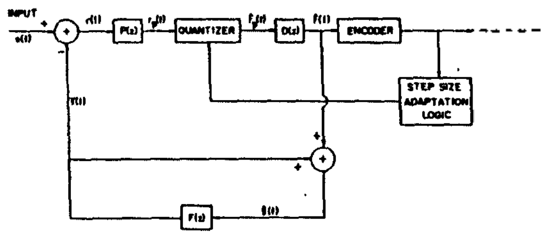

Fig. 3. Block diagram of a modified ADPCM-PPF transmitter.

errors, the output signal of the ADPCM-PPF is expressed as (5), and the input to the quantizer is given by

$$R_p(z) = \{S(z) - \frac{\hat{R}(z)}{1 - F(z)}F(z)\}P(z), \tag{7}$$

where $R_p(z)$ is the z-transform of pre-filtered residual signal $r_p(t)$. Also, the quantization noise Q(z) is expressed as

$$Q(z) = \hat{R}_p(z) - R_p(z), \tag{8}$$

and R(z) is given by

$$\hat{R}(z) = \hat{R}_p(z) \cdot D(z). \tag{9}$$

Using (5), (7), (8) and (9), we obtain

$$\hat{S}(z) = S(z) + \frac{Q(z)}{P(z)} \tag{10}$$

We can note that (10) corresponds to (4) since

$$D(z) = \frac{1}{P(z)}.$$

As for ADM, the basic structure of an ADM system with pre- and post-filters (ADM-PPF) is the same as that shown in Fig. 1 with the ADPCM replaced by an ADM coder. For ADM, we use continuously variable slope delta modulation (CVSD) in our study [14]. As in the ADPCM, the noise shaping effect of quantization noise by using the pre- and post-filters in ADM is such that the quantization noise in the perceptually important frequency region of 1 to 3 kHz can be reduced.

III. PERFORMANCE OF A WAVEFORM QUANTIZER WITH FWMSE AS A CRITERION

Previously, optimum preemphasis and deemphasis filters that minimize the mean-squared error (MSE) was investigated [15]. In this work, we consider optimum pre- and post-filters that minimize the FWMSE in speech waveform coding. A waveform quantizer with pre- and post-filters under consideration is shown in Fig. 4(a). In general, the quantization noise of a waveform coder can be represented approximately by an additive white noise model [16]. This is shown in Fig. 4(b).

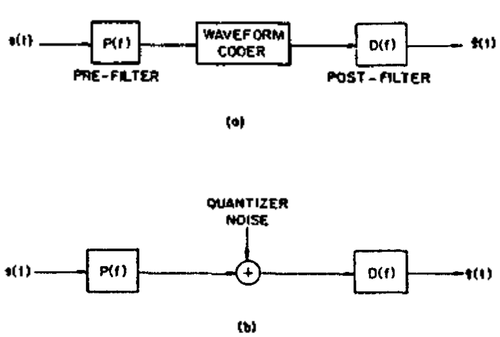

Fig. 4. (a) A waveform coder with pre- and post-filters (b) Corresponding model.

Let the input signal s(t) be a stationary process with spectral density S(f) and bandlimited to B Hz. Here, we wish to determine the transfer functions of the pre-filter P(f) and the post-filter D(f) that minimize the FWMSE.

We assume that P(f) is normalized such that the average output power with input s(t) is unity. This power constraint is expressed as

$$\int_0^B |P(f)|^2\, S(f)df = 1. \qquad (11)$$

Also, we assume that the post-filter D(f) is inversely proportional to the pre-filter P(f) in reconstructing the original signal. This is,

$$D(f) = \frac{1}{P(f)}. \qquad (12)$$

Then, the FWMSE may be written as

$$\overline{e_W^2(t)} \triangleq \int_0^B NW(f)|D(f)|^2\, df \qquad (13)$$

where N is the spectral density of quantization noise and W(f) is a frequency weighting function. From (11), (12) and (13), and using an undetermined multiplier α, we have

$$\frac{\partial}{\partial P(f)}\left\{\int_0^B NW(f)\frac{1}{|P(f)|^2}df + \alpha\int_0^B |P(f)|^2 S(f)df\right\} = 0 \qquad (14)$$

Then, from (11) and (14), the optimum $\dot{P}(f)$ that yields the minimum error $e_W^2(t)_{min}$ is given by

$$|P(f)|^4 = k\cdot\frac{W(f)}{S(f)} \qquad (15)$$

where

$$k \triangleq \frac{1}{\left\{\int_0^B W(f)^{\frac{1}{2}}S(f)^{\frac{1}{2}}df\right\}^2}.$$

Eq. (15) shows that the pre-filter minimizing the FWMSE must be designed such that its transfer function is proportional to the fourth root of the frequency weighting function used, but inversely proportional to the fourth root of the spectral amplitude of input speech.

We note that the FWSNR of a conventional waveform coder without the pre- and post-filters can be written as

$$\mathrm{FWSNR}\Big|_{\text{without PPF}} = 10\log\frac{\int_0^B S(f)df}{\int_0^B NW(f)df}. \tag{16}$$

As for a waveform coder with the pre- and post-filters, the FWSNR is given by

$$\mathrm{FWSNR}\Big|_{\text{with PPF}} = 10\log\frac{\int_0^B S(f)df}{\int_0^B NW(f)|D(f)|^2 df}. \tag{17}$$

From (16) and (17), we see that the use of frequency-weighted pre- and post-filters yields a performance improvement in FWSNR equal to

$$\eta = 10\log\frac{\int_0^B W(f)df}{\int_0^B W(f)|D(f)|^2 df}\,\mathrm{dB}. \tag{18}$$

One can note from (12), (15) and (18) that in order to have a performance gain (i.e., $\eta > 0$) the post-filter must be designed such that the spectral magnitude of $|D(f)|^2$ is less than 1 for the frequency range of 0 to B Hz. If we use as a model of real speech the first-order Gauss-Markov process with correlation ρ and its spectrum given by

$$S(f) = \frac{1}{1-2\rho\cos 2\pi f+\rho^2}, \tag{19}$$

and specify a weighting function (e.g., the C-message weighting function), we can compute approximately the performance gain resulting from the use of pre- and post-filters by using (12), (15) and (18).

IV. COMPUTER SIMULATION RESULTS

In our simulation, we used for noise spectral shaping the pre- and postfilters that minimize MSE or FWMSE. Unlike the conventional noise shaping filter which is in general complex, the pre- and post-filters shown in Fig. 5 are relatively simple. They were used in ADPCM and ADM to improve the subjective quality. In general,

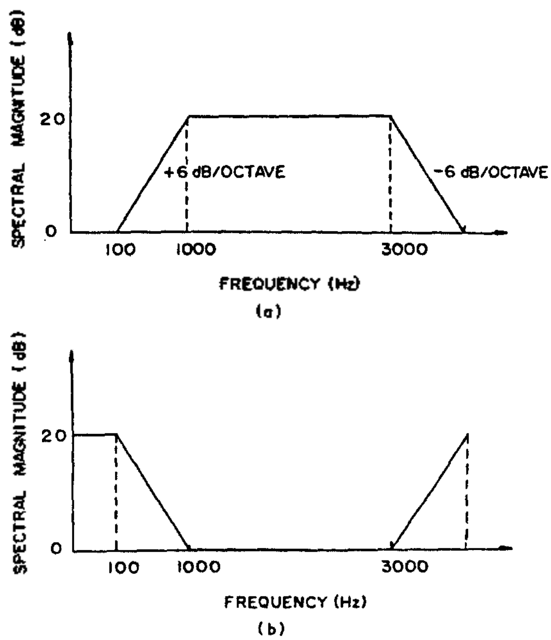

Fig. 5. Frequency responses of (a) pre- and (b) post-filters.

the pre- and post-filters minimizing MSE are realized as simple preemphasis and deemphasis filters that are designed based on the spectral envelope of speech. On the other hand, the pre- and post-filters minimizing the FWMSE are based on the spectral envelope of speech and also on a frequency weighting functing used as discussed in Section III.

It is known that the segmental frequency-weighted SNR($FWSNR_{SEG}$) gives best correlation with the subjective quality of a speech waveform coder [17],[18]. In this work, we used the C-message weighting function to obtain the frequency-weighted noise power. The segmental FWSNR which we used as a performance criterion is defined as

$$FWSNR_{SEG}(dB) = \frac{10}{N} \sum_{i=1}^{N} \log_{10}(1 + FWSNR_i) \tag{20}$$

where

$$FWSNR = \frac{\sum_{k=1}^{N} \sum_{\ell=1}^{M} S_{k,\ell}^2(f)}{\sum_{k=1}^{N} \sum_{\ell=1}^{M} E_{k,\ell}^2(f) \cdot W_{\ell}^2(f)}, \tag{21}$$

N is the number of frames of input speech, and M is the number of frequency components in one frame. The procedure for obtaining the FWSNR is shown in Fig. 6. In this figure, $s_{k,\ell}$ is the ℓ-th input speech sample of the k-th block and $\hat{s}_{k,\ell}$ is its reproduction. The error sequence $\{e_{k,\ell}\}$ is the difference of the sequences $\{s_{k,\ell}\}$ and $\{\hat{s}_{k,\ell}\}$. Here $S_{k,\ell}(f)$ is the spectral-weighted magnitude of ℓ-th frequency component in the k-th input signal block and $E^W_{k,\ell}(f)$ is the spectral-weighted magnitude of the ℓ-th frequency component in the k-th error signal block.

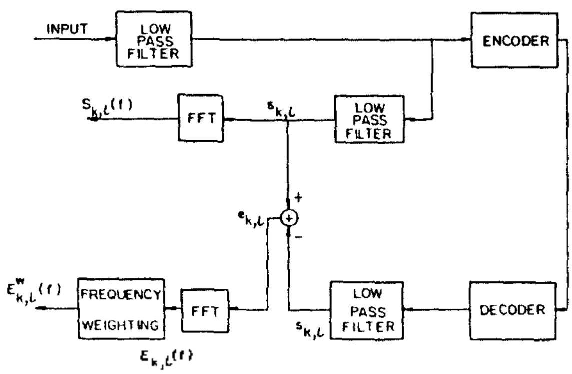

Fig. 6. Procedure for computing frequency-weighted quantization noise.

Simulations of the waveform coders (ADPCM and ADM) were performed with real speech signals bandlimited to 3.4 kHz and sampled at 8 and 16 kHz. The waveform coders under study were simulated with a step size range of 60 dB and the input signal level varied over a range of 80 dB.

The performance of ADPCM is shown in Fig. 7. This figure shows the $FWSNR_{SEG}$ of ADPCM with and without the pre- and post-filters. It is seen that, when we use the pre- and post-filters in the ADPCM with the Jayant's quantizer [13] and a fixed one-tap predictor, the performance can be improved by about 3 dB. Also, it is noted that the performance resulting from the use of the pre- and post-filters that minimize the FWMSE is superior to that resulting from the use of pre- emphasis and de-

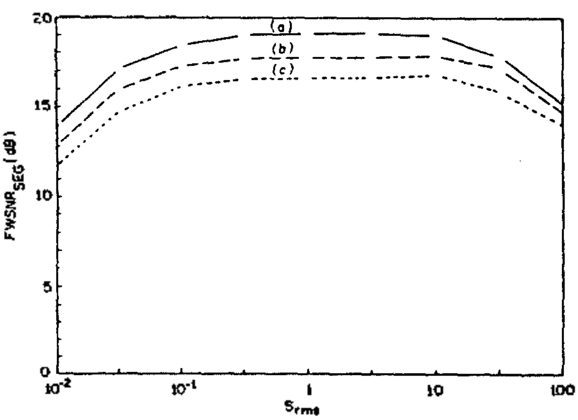

Fig. 7. $FWSNR_{SEG}$ of ADPCM with a 1-tap fixed predictor vs. input signal level (b=2).
(a) With pre- and post-filtering
(b) With preemphasis and deemphasis
(c) Without preprocessing

emphasis filters minimizing the MSE. In addition, Figs. 8 and 9 show noise spectral shaping by the pre- and post-filters for ADPCM with and without adaptive prediction, respectively. We can note from these figures that the quantization noise in the frequency range of 0.5 to 3 kHz is reduced as a result of noise spectral shaping by the pre- and post-filters.

As for ADM, a peak performance gain of about 2 to 3 dB in $FWSNR_{SEG}$ could be obtained. The result is shown in Fig. 10. The performance gain resulting from the use of pre- and post-filtering becomes larger when the input signal level is lower. The reason may be due to the fact that noise spectral shaping is effective in the granular noise region. Fig. 11 shows the spectra of the original speech and quantization noise of CVSD and CVSD-PPF at the rate of 16kbits/s.The effectiveness of noise spectral shaping can be seen in the figure.Clearly,we can see that the noise in the frequency region of 1 to 3 kHz has been reduced. We also note from Fig. 11 that the pre- and post-filters minimizing the FWMSE is more effective for real speech than the pre- and post-filters minimizing the MSE.

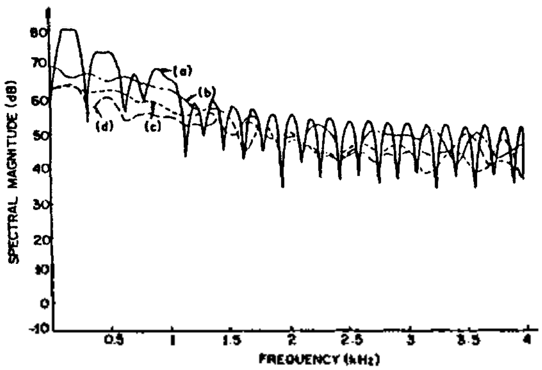

Fig. 8. Spectral envelopes of input speech and output quantization noise with and without noise spectral shaping for ADPCM with a 1-tap fixed predictor (b=2).
(a) Original speech
(b) Output quantization noise
(c) Output quantization noise shaped by pre-emphasis and deemphasis filters.
(d) Output quantization noise shaped by pre- and post-filters.

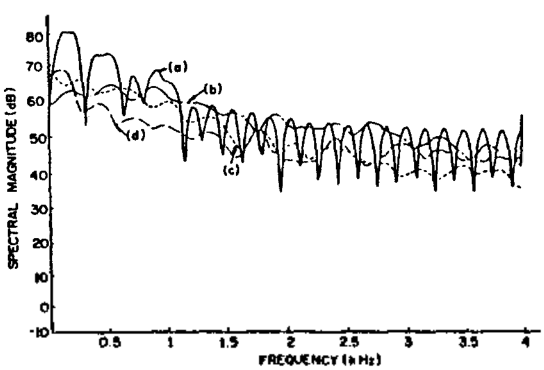

Fig. 9. Spectral envelopes of input speech and output quantization noise with and without noise spectral shaping for ADPCM with adaptive predictor.
(a) Original speech
(b) Output quantization noise
(c) Output quantization noise shaped by pre-emphasis and deemphasis filters.
(d) Output quantization noise shaped by pre- and post-filters.

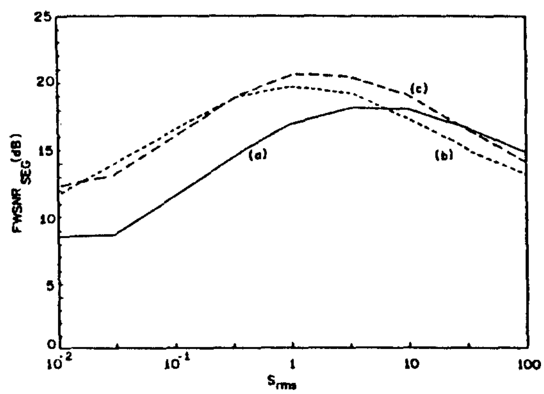

Fig. 10. $FWSNR_{SEG}$ of CVSD vs. input signal level (f_s=16 kHz).
(a) Without preprocessing
(b) With preemphasis and deemphasis
(c) With pre- and post-filters.

In general, quantization noise of a waveform coder with syllabic companding is largely granular when the signal level is low. Also, it is known that granular noise is subjectively more annoying than overload noise [19]. Hence, the use of pre- and post-filters for noise spectral shaping in the granular noise region should result in perceptually pleasing sound.

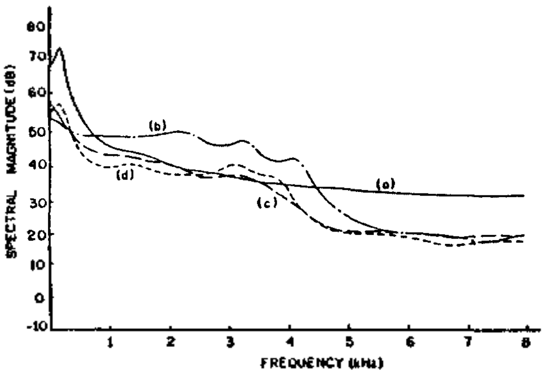

Fig. 11. Spectral envelopes of input speech and output quantization noise with and without noise spectral shaping for CVSD (f_s=16 kHz).
(a) Original speech
(b) Output quantization noise
(c) Output quantization noise shaped by pre-emphasis and deemphasis filters.
(d) Output quantization noise shaped by pre- and post-filters

V. CONCLUSIONS

The use of pre- and post-filters in waveform coders has been studied for improvement of the speech quality. The performance gain due to the adaptive pre-and post-filtering has been obtained analytically for real speech. Also, the noise shaping effect by various pre- and post-filters has been shown and discussed. According to our simulation results, when the pre- and post-filters are used in ADPCM with a fixed or adaptive predictor, the performance can be improved by about 3 dB in $FWSNR_{SEG}$. In general, the performance resulting from the use of the pre- and post-filters that minimize the FWMSE is superior to that resulting from the use of pre-emphasis and deemphasis filters minimizing the MSE. As for ADM, a similar performance gain can be achieved. The performance gain resulting from the use of pre-and post-filtering becomes larger when the input signal level is lower. In addition, it has been shown that the quantization noise in the perceptually important frequency region of 1 to 3 kHz can be reduced when the pre- and post-filters are used in ADPCM and ADM.

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, May 1982.
3. H.S. Lee and C.K. Un, "On the performance of speech waveform coders with noise spectral shaping," IEEE Trans. Commun., vol. COM-33, pp.742-746, July 1985.
4. D. Chan and R.W. Donaldson, "Subjective evaluation of pre- and post-filtering in PAM, PCM, and DPCM voice communication systems," IEEE Trans. Commun. Technol., vol. COM-19, pp.188-199, Oct. 1971.
5. J.D. Markel and A.H. Gray, Jr., Linear Prediction of Speech, Springer-Verlog, Berlin, 1976.
6. P. Noll, "On predictive quantizing schemes," Bell Syst. Tech. J., vol. 57, pp.1499-1532, Mar-June 1978.
7. C.C. Cutler, "Differential quantization of communications," U.S. Patant 2-605-361, 1952.

8. 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.
9. B.S. Atal and M.R. Schroeder, "Predictive coding of speech signals and subjective error cirteria," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, pp.247-254, June 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, Mar.-Apr. 1981.
12. A.J. Aikens and D.A. Lewinski, "Evaluation of message circuit noise," Bell Syst. Tech. J., vol. 39, pp.879-910, July 1960.
13. P. Cummiskey, N.S. Jayant, and J.L. Flanagan, "Adaptive quantization in differential PCM coding of speech," Bell Syst. Tech. J., vol. 52, pp.1105-1118, Sept. 1973.
14. 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.
15. J.P. Costas, "Coding with linear systems," Proc. IRE, vol. 40, pp.1101-1103, Sept. 1952.
16. B. Widrow, "A study of rough amplitude quantization by means of the Nyquist sampling theorem," IRE Trans. Circuit Theory, vol CT-3, pp.266-276, Dec. 1956.
17. 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.
18. T.P. Barnwell, III, "Correlation analysis of subjective and objective measures for speech quality," in Proc. 1980 IEEE Int. Conf. Acoust., Speech, Signal Processing, pp.706-709, Apr. 1980.
19. H. Levitt, C.A. McGonegal and L.L. Cherry, "Perception of slope-overload distortion in delta-modulated speech signals," IEEE Trans. Audio Electroacoust., vol. AU-18, pp.240-247, Sept. 1970.