AT&T Bell Laboratories Technical Journal
Vol. 63, No. 8, October 1984
Printed in U.S.A.

Enhancement of ADPCM Speech by Adaptive Postfiltering

By V. RAMAMOORTHY* and N. S. JAYANT†

(Manuscript received February 8, 1984)

Adaptive Differential Pulse Code Modulation (ADPCM) systems can provide high-quality digitizations of telephone-bandwidth speech at a bit rate of 32 kb/s. At a lower bit rate such as 24 kb/s, the quality of the speech is limited by an easily perceptible level of quantization noise. This paper proposes an adaptive postfiltering procedure that can significantly enhance the quality of lower bit rate ADPCM. The coefficients of the postfilter are easily derivable from the predictor coefficients in the ADPCM decoder. In a subjective test involving 18 listeners and two sentence-length test inputs, the enhanced 24-kb/s speech with an optimized postfilter design ranks very close to conventional 32-kb/s speech. A suggested application of the postfiltering procedure is in packet voice or mobile radio systems where substandard bit rates such as 24 kb/s or 16 kb/s are sometimes necessary. The postfiltering algorithm has also been successfully tested in non-DPCM situations, such as in the enhancement of speech degraded by additive white Gaussian noise.

I. INTRODUCTION

Recent algorithms for adaptive prediction[1] and adaptive quantization[2] have led to the realization of high-quality ADPCM systems at 32 kb/s. This bit rate is the result of 8-kHz sampling and quantization using 4 bits/sample. The quality of 24-kb/s speech using the same prediction algorithm and 3-bits/sample coding is limited by a clearly perceptible level of quantization noise. This paper proposes a very simply implemented postfiltering algorithm, which provides a significant enhancement of 24-kb/s quality. In a subjective test to be

* University of Linköping, Sweden. † AT&T Bell Laboratories.

described later in this paper, the enhanced 24-kb/s ADPCM system was ranked very close to conventional 32-kb/s ADPCM.

A natural application of the postfiltering procedure would be in variable bit rate ADPCM systems such as packet networks or mobile radio where substandard bit rates such as 3 or 2 bits/sample are occasionally encountered. The postfiltering technique described in this paper is particularly effective at the bit rate of 3 bits/sample. It is also effective in non-DPCM situations such as in the enhancement of speech degraded by additive white Gaussian noise. When the signal-to-noise ratio (s/n) at the input to the postfilter is too low (as in 2-bit ADPCM or with white Gaussian noise at a relative noise level exceeding approximately −3 dB), noise suppression can only be achieved at the expense of severe distortion of the speech signal itself. When the 24-kb/s ADPCM is enhanced, the introduction of speech distortion is perceptible, but the effect of noise reduction is by far the more dominant phenomenon.

II. A SEMIQUANTITATIVE EXPLANATION OF THE POSTFILTERING TECHNIQUE

The philosophy of the postfiltering technique is represented in Fig. 1. Part (a) of the figure shows a signal spectrum with two narrowband components in the frequency regions W_1 and W_2, and a flat noise spectrum that is 15 dB below the first signal component but 5 dB above the second signal component. An ideal postfilter for this situation would have a gain of unity (0 dB) in the regions W_1 and W_2 and a gain of zero ($-\infty$ dB) in the rest of the frequency range. In real speech applications, implementation of such all-or-none responses is impractical except in the special cases where the stopband regions of the postfilter are merged into a single contiguous frequency region as in a low-pass or high-pass postfilter.[3,4]

A more practical approach, proposed in this paper, is the use of a postfilter frequency response that peaks in the regions W_1 and W_2, but is significantly lower in the rest of the frequency range. Figure 1b illustrates an extreme example of this approach. Here, the transfer function of the postfilter is chosen to be identical to the input signal spectrum in Fig. 1a. The resulting spectra of postfiltered signal and postfiltered noise preserve the original signal-to-noise ratios of 15 dB and −5 dB in the regions W_1 and W_2, respectively. However, the noise in the rest of the illustrated frequency range is now much lower, relative to the signal levels, than in part (a) of the figure. Specifically, the signal-to-background-noise ratios for regions W_1 and W_2 are now 45 dB and 10 dB, in place of 15 dB and −5 dB in the absence of postfiltering. This suppression of background noise also implies that

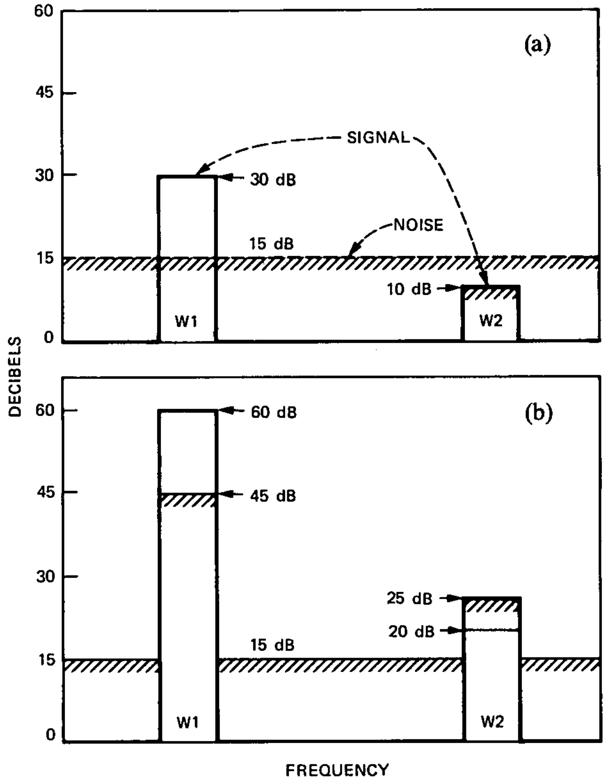

Fig. 1—An idealized explanation of the effects of postfiltering, assuming a signal with two narrowband components and a noise spectrum that is white. (a) Signal and noise spectra at the input to the postfilter, showing signal-to-noise ratios of 15 dB and −5 dB in signal frequency bands W_1 and W_2. (b) Spectra of postfiltered signal and postfiltered noise, assuming a postfilter transfer function identical to the signal spectrum in (a). Regions W_1 and W_2 continue to have local signal-to-noise ratios of 15 dB and −5 dB as in (a), but the signals are now 45 dB and 10 dB above the out-of-band noise level. In (a) the corresponding numbers are only 15 dB and −5 dB. The overall effect is a reduction of perceived noise, but the price paid is a change in the relative strengths of the signal components in W_1 and W_2.

the residual noise spectrum after postfiltering is very similar to the input signal spectrum itself. In speech applications, noise that is shaped in this manner tends to be perceived as speech.

A postfiltering operation such as that in Fig. 1b provides a significant amount of signal enhancement for the reasons just described. It should be noted, however, that such a postfilter also distorts the signal. For example, the difference in signal levels in the regions W_1 and W_2 has been distorted, from 20 dB in Fig. 1a to 35 dB in Fig. 1b. The postfiltering technique to be described in the next section provides a controlled exchange between signal distortion and noise suppression.

In the applications discussed, the technique realizes a broad range of postfilter design over which the phenomenon of noise suppression dominates the phenomenon of signal distortion.

Although speech enhancement[5] is an "ancient" art, we believe that the adaptive postfiltering technique discussed in the next section is novel. It can be used as a very general technique for speech enhancement. It can also be used very effectively in the specific context of ADPCM noise. The coefficients of the proposed postfilter are inspired by the coefficients of the adaptive predictor in ADPCM coding, and are in fact very closely related to these coefficients.

III. POSTFILTERING OF ADPCM SPEECH

Figures 2 and 3 provide block diagram descriptions of ADPCM with adaptive postfiltering.

Figure 2 shows the decoder part of the system. Broken lines in the figure refer to parts of the system that compute the coefficients of the adaptive predictor and the adaptive postfilter. The coefficients used in the postfilter are differently scaled versions of the coefficients used in the adaptive predictor. These coefficients are already available in conventional ADPCM. In the case of a system with Backward-Adaptive Prediction (APB), the predictor coefficients are updated in gradient-search algorithms driven by a recent history of the input and output of the ADPCM decoder.

A more complete block diagram of the ADPCM system appears in Fig. 3. The quantizer and predictor assumed in this paper are both

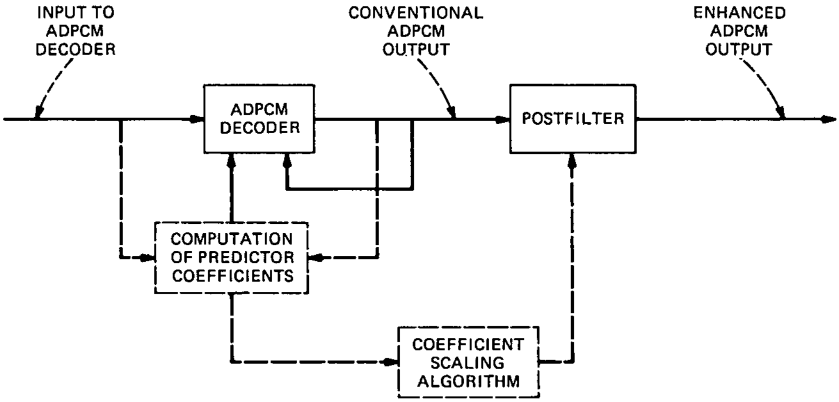

Fig. 2—Adaptive postfiltering of the output of an ADPCM decoder. The coefficients of the postfilter are scaled versions of the coefficients of the adaptive predictor in DPCM. In DPCM-APB, the predictor coefficients are obtained on the basis of observations of a recent history of decoder input and decoder output. The parts of the circuit that determine coefficient values are shown by broken lines.

backward-adaptive devices, implying that no special side information needs to be explicitly transmitted to the ADPCM decoder to enable adaptations of quantizer step size and predictor coefficients.

The adaptive quantizer assumed in this paper is one based on the use of a one-word memory,[6] but the results of this paper are fully expected to extend to a system that may use the more generalized version of this quantizer, as described in Ref. 2. In the quantizer used in this paper, the ratio of maximum step size to minimum step size is 512, and the minimum step size is in the order of 2^{-12} times the peak-to-peak value of the speech input $x(n)$.

This adaptive predictor we assumed is a pole-zero predictor, similar to that in Ref. 1. As Fig. (3) shows, the predicted value $\hat{x}(n)$ of input $x(n)$ is a combination of two components, the outputs $\hat{x}_z(n)$ and $\hat{x}_p(n)$ of an all-zero predictor $B(z)$ and an all-pole predictor $A(z)$. Formally,

$$\hat{x}(n) = \hat{x}_z(n) + \hat{x}_p(n)$$

$$\hat{x}_p(n) = \sum_{j=1}^{2} a_j(n)y(n-j)$$

$$\hat{x}_z(n) = \sum_{j=1}^{6} b_j(n)u(n-j),$$

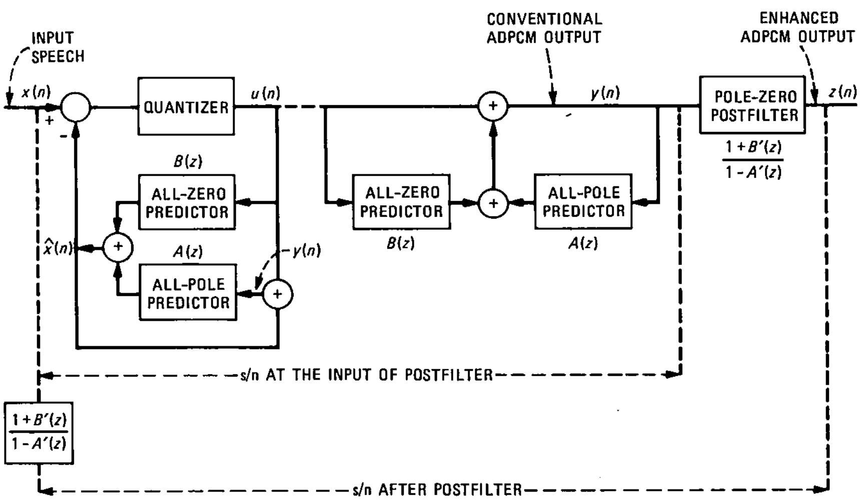

Fig. 3—Complete block diagram of an ADPCM system with a pole-zero predictor [defined by all-zero and all-pole components $A(z)$ and $B(z)$] and a pole-zero postfilter [defined by components $A'(z)$ and $B'(z)$ that are derived from $A(z)$ and $B(z)$]. The extreme case of $A'(z) = B'(z) = 0$ results in conventional ADPCM without postfiltering. The case of $A'(z) = A(z)$ and $B'(z) = B(z)$ results in a postfilter transfer function that is identical to the input signal spectrum, as in Fig. 1b.

where $u(n)$ is the quantized version $Q[\cdot]$ of the prediction error and $y(n)$ is the reconstructed output:

$$u(n) = Q[x(n) - \hat{x}(n)]$$
$$y(n) = \hat{x}(n) + u(n).$$

Adaptation of the predictor coefficients a_j and b_j follow the updating algorithms[7]

$$a_j(n) = \lambda_j a_j(n-1) + \mu_j \mathrm{sgn}[u(n-1)]\mathrm{sgn}[y(n-1-j)]$$
$$j = 1, 2; \qquad \lambda_1 = 511/512; \qquad \lambda_2 = 255/256; \qquad \mu_1 = \mu_2 = 0.008$$

and

$$b_j(n) = \lambda_j' b_j(n-1) + \mu_j' \mathrm{sgn}[u(n-1)]\mathrm{sgn}[u(n-1-j)]$$
$$j = 1 \text{ to } 6; \qquad \lambda_j' = 255/256 \quad \text{and} \quad \mu_j' = 0.008 \quad \text{for all } j.$$

The coefficients of the all-pole predictor are further controlled, for stability reasons, by the following constraints:

$$-0.75 \leq a_2 \leq 0.97$$
$$|a_{1,\max}| = 0.97 - a_2; \qquad |a_1| = \min\{|a_1|, |a_{1,\max}|\}$$
$$a_1 = |a_1| \mathrm{sgn}\, a_1.$$

3.1 Coefficients of the postfilter

A good starting point for designing the postfilter is the frequency response of the inverse predictor. This is the system whose input and output are the innovations $u(n)$ and the reconstruction $y(n)$. Its transfer function, derivable from linear equations that relate $u(n)$, $\hat{x}(n)$, and $y(n)$, is

$$\frac{Y(z)}{U(z)} = \frac{1 + B(z)}{1 - A(z)},$$

where

$$A(z) = \sum_{j=1}^{2} a_j z^{-j}; \qquad B(z) = \sum_{j=1}^{6} b_j z^{-j}.$$

The speech-like transfer function of Fig. 1b is approximated if the postfilter response is identical to the function $[Y(z)]/[U(z)]$. This is because the spectrum of the quantized innovations $u(n)$ is approximately white and that of the reconstruction $y(n)$ is hopefully an approximation to that of the input $x(n)$. More generally, as in Fig. 3, we propose a postfilter transfer function

$$F(z) = \frac{1 + B'(z)}{1 - A'(z)},$$

where

$$A'(z) = \sum_{j=1}^{2} a_j \alpha^j z^{-j}; \qquad B'(z) = \sum_{j=1}^{6} b_j \beta^j z^{-j}$$

$$0 \leqslant \alpha \leqslant 1; \quad \text{and} \quad 0 \leqslant \beta \leqslant 1.$$

The extreme situation of Fig. 1b is approximated when $\alpha = \beta = 1$. In practice, this approximation can be quite poor because of the effects of nonideal predictor adaptation, usually resulting in an inverse predictor transfer function that is a flattened version of the input speech spectrum, with poles and zeros that may also be significantly shifted from their original locations. The case of $\alpha = \beta = 0$ corresponds to conventional ADPCM without any postfiltering. As we discuss in the next section, intermediate designs provide different mixes of noise suppression and speech distortion.

Figure 4 shows an illustrative spectrum of input speech and compares it with the transfer functions $F(z)$ for ($\alpha = 0.2$; $\beta = 1.0$) and ($\alpha = 1.0$; $\beta = 1.0$). The latter condition simply corresponds to the transfer function of the inverse predictor.

IV. EXPERIMENTAL RESULTS WITH ADPCM SPEECH

The speech inputs used in the experiment were the sentence-length

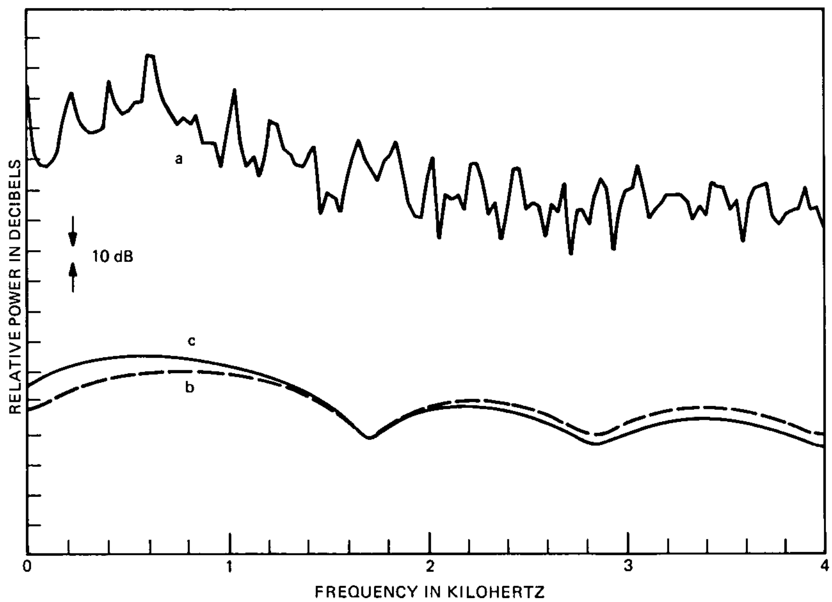

Fig. 4—(a) Input speech spectrum; and power transfer functions of postfilter with scaling coefficients for (b) $\alpha = 0.2$, $\beta = 1.0$, and for (c) $\alpha = 1.0$, $\beta = 1.0$. The plot (c) is merely the transfer function of the inverse predictor in the DPCM-APB system. [The 0-dB line is the same for (b) and (c) but different for (a)].

utterances “The Lathe is a big tool” and “The chairman cast three votes,” bandlimited to 3.2 kHz in each case and sampled at 8 kHz. These inputs will be referred to as L8 and C8, respectively.

4.1 Signal-to-noise ratio results

Figures 1a and 1b indicate that postfiltering can result in significant improvements in s/n. Table I further demonstrates this for the examples of 3-bit and 2-bit DPCM. The results tabulated are the values of the s/n at the input of the postfilter and the s/n after postfiltering. (See Fig. 3.) Table I also shows corresponding values of the segmental s/n. In the ranges $0 \leqslant \alpha \leqslant 1$ and $0 \leqslant \beta \leqslant 1$ for the coefficient scaling factors, the greatest gains in the s/n are obtained when $\alpha = \beta = 1$. These gains are seen to be as high as 8.9 dB for both L8 and C8. The gains of the s/n at the input of the postfilter are always lower for the design $\alpha = 0.2$ and $\beta = 1.0$. But we presently note that these settings of α and β provide a subjectively desirable design.

4.2 Subjective results

Tables II and III provide the results of a subjective test involving 14 listeners, including 9 from the AT&T Bell Laboratories Acoustics Research department and 5 listeners who had no prior exposure to speech coding experimentation or testing. A total of eight stimuli were included in the test. These included 4-bit ADPCM without postfiltering, 3-bit ADPCM with six postfiltering conditions (including the no-postfiltering case of $\alpha = \beta = 0$), and 4-bit ADPCM with 6-kHz sampling and a substandard speech bandwith of $W = 2.6$ kHz. This last condition was included to provide a 4-bit, 24-kb/s alternative to the 3-bit ADPCM stimuli, all of which also had a bit rate of 24 kb/s. The values of α and β used in the test were selected on the basis of a pilot test

Table I—Values of s/n at input of postfilter and after postfiltering (see Fig. 3). Numbers in parentheses are corresponding values of segmental s/n ratio

Input	R (bits/ sample)	s/n at Input of Postfiltering	s/n After Postfiltering (dB) $\alpha = 1.0$ $\beta = 1.0$	s/n After Postfiltering (dB) $\alpha = 0.2$ $\beta = 1.0$
L8	3	21.7 (23.6)	28.9 (32.5)	27.0 (28.1)
	2	15.2 (17.1)	20.9 (25.3)	19.5 (21.0)
C8	3	18.9 (20.7)	27.2 (29.6)	23.9 (24.6)
	2	12.7 (14.9)	20.0 (22.9)	17.1 (18.2)

Table II—Number of wins in a round-robin tournament involving eight coding conditions and four listeners, where the maximum possible score is 196 for any given coding condition

Bits/ Sample	4	3						4 ($W =$ 2.6 kHz)
α, β	0, 0	0, 0	0.2, 1.0	0.4, 0.8	0.4, 0.6	0.6, 0.6	0.6, 0.4	0, 0
L8	127	75	118	116	104	113	106	25
C8	111	80	129	120	106	117	102	19

Table III—Rank ordering of coding conditions by the group of 14 listeners and by a subgroup of 9 listeners from the Acoustics Research Department

Bits/ Sample	4	3						4 ($W =$ 2.6 kHz)
α, β	0, 0	0, 0	0.2, 1.0	0.4, 0.8	0.4, 0.6	0.6, 0.6	0.6, 0.4	0, 0
L8 (G14)*	1	7	2	3	6	4	5	8
L8 (G9)†	1	7	3	2	6	4	5	8
C8 (G14)	4	7	1	2	5	3	6	8
C8 (G9)	2	7	1	3	5	4	6	8

* G14 group of 14 listeners.
† G9 group of 9 listeners.

that identified the interesting ranges of these parameters from the point of view of perceived mixes of noise suppression and speech distortion.

In general, use of postfiltering results in an amplification of the speech signal as suggested in Fig. 1b. The postfiltered speech stimuli were therefore appropriately scaled down to mitigate differences in stimulus loudness.

The subjective test involved an exhaustive pairwise comparison of all possible stimulus pairs, with each pair appearing at random places in the test once in each possible order of presentation. The total number of AB comparisons was therefore 768 ($8 \times 7 = 56$ possible stimulus pairs for each of 14 listeners).

Table II shows, separately for inputs L8 and C8, the total number of wins of each stimulus, with a maximum possible score of 196 for each stimulus [a maximum score of $2 \cdot (8 - 1)$ for each of 14 listeners]. It is seen that the worst two coding conditions stand apart from the rest. These conditions are 3-bit ADPCM with no prefiltering and 4-bit ADPCM with 2.6-kHz bandwidth speech input (and output). This latter condition gets a particularly low total score. Table II also shows that the above results are not very different for the inputs L8 and C8.

Table III shows, separately for inputs L8 and C8, the rank ordering of the eight coding conditions in the subjective test. Results are shown separately for the total group of 14 listeners and the group of 9 listeners from the Acoustics Research department. It is seen that the rankings are not significantly different for the two populations. The best setting of the coefficient scaling parameters is defined by

$$\alpha = 0.2; \qquad \beta = 1.0$$

in each case, and for both L8 and C8. With input L8, 3-bit ADPCM postfiltered as above is ranked a close second to 4-bit ADPCM speech of equal bandwidth. In fact, the 4-bit ADPCM coder is ranked only fourth when the results of all 14 listeners are pooled together. The second and third ranks in this category belong to postfilters with the design ($\alpha = 0.4$; $\beta = 0.8$) and ($\alpha = 0.6$; $\beta = 0.6$). The preference for the design ($\alpha = 0.2$; $\beta = 1.0$) has a simple interpretation. It suggests a postfilter transfer function that mimics the approximate speech spectrum (the inverse predictor function) very closely at the zeros of that spectrum ($\beta = 1.0$), but very loosely at the poles ($\alpha = 0.2$). This suggests a condition that seeks to maximize background noise suppression and minimized perceived speech distortion. In the case of voiced speech segments, the poles tend to correspond to formant frequencies and the value of $\alpha = 0.2$ prevents an undue emphasis of the higher-amplitude spectral peaks, a situation that was indeed encountered in the example of Fig. 1b.

4.3 Enhancement of 2-bit ADPCM

As we see in Table I, the s/n gains due to postfiltering are equally significant for both 3-bit ADPCM and 2-bit ADPCM. Perceptually, however, the general noise level in 2-bit ADPCM speech is such that a useful degree of noise suppression requires the design of ($\alpha = 1.0$; $\beta = 1.0$). With this design, the speech distortion introduced by the postfilter is also substantial. For this reason, the case of 2-bit ADPCM is not considered to be of sufficient practical importance to pursue formal subjective testing. Informal testing shows, however, that the design of ($\alpha = 1.0$; $\beta = 1.0$) is again preferable to conventional 2-bit ADPCM ($\alpha = 0$; $\beta = 0$).

V. ENHANCEMENT OF SPEECH DEGRADED BY ADDITIVE WHITE GAUSSIAN NOISE

The specific postfiltering algorithm of Fig. 1b ($\alpha = 1.0$; $\beta = 1.0$) was also applied to speech degraded by additive white Gaussian noise. The input speech was L8, the speech-to-noise ratios ranged from −3 dB to 17 dB, and all coefficients were obtained simply by simulating the easily available case of 5-bit ADPCM, a bit rate high enough to introduce very little quantization noise in comparison with the levels of Gaussian noise being studied. We find the postfiltering algorithm provides a very useful enhancement of noisy speech if the input to the postfilter had a s/n of at least +3 dB. For lower values of s/n, postfiltering provides noise suppression, but at the cost of substantial distortion of the speech itself.

REFERENCES

1. T. Nishitani et al., "A 32 kb/s Toll Quality ADPCM Codec Using a Single Chip Signal Processor," Proc. ICASSP (April 1982), pp. 960–3.
2. D. W. Petr, "32 kb/s ADPCM-DLQ Coding for Network Applications," Proc. IEEE Globecom Conf., December 1982. pp. A8.3.1–A8.3.5.
3. N. S. Jayant, "Adaptive Postfiltering of ADPCM Speech," B.S.T.J., *60*, No. 5 (May–June 1981), pp. 707–17.
4. J. O. Smith and J. B. Allen, "Variable Bandwidth Adaptive Delta Modulation," B.S.T.J., *60*, No. 5 (May–June 1981), pp. 719–37.
5. J. S. Lim (ed.), *Speech Enhancement*, Englewood Cliffs, NJ: Prentice Hall, 1983.
6. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," B.S.T.J., *52*, No. 7 (September 1973), pp. 1105–18.
7. D. W. Petr, unpublished work.

AUTHORS

Nuggehally S. Jayant, B.Sc. (Physics and Mathematics), Mysore University, 1962, B.E., 1965, and Ph.D. (Electrical Communication Engineering), 1970, Indian Institute of Science, Bangalore; Research Associate, Stanford University, 1967–1978; AT&T Bell Laboratories, 1968—. Mr. Jayant was a visiting scientist at the Indian Institute of Science in 1972 and 1975 and a Visiting Professor at the University of California, Santa Barbara, in 1983. Mr. Jayant has worked in the field of digital coding and transmission of waveforms, with special reference to robust speech communications. Editor, IEEE Reprint Book, *Waveform Quantization and Coding* and co-author of *Digital Coding of Waveforms: Principles and Applications to Speech and Video* (Prentice Hall, 1984).

Venkatasubbarao Ramamoorthy, B.E. (Electrical Engineering), 1970, The Regional Engineering College at Tiruchirappalli, India; M.Tech., 1972, The Indian Institute of Technology, Madras, India; Tekn.Dr., 1981, University of Linköping, Linköping, Sweden. Mr. Ramamoorthy was with the Indian Space Research Organisation at Bangalore, India, prior to his joining as a staff member at the department of Electrical Engineering, the University of Linköping, Sweden, in 1974. He visited AT&T Bell Laboratories during the summer of 1983. His current research interests include speech processing in mobile and packet radio environments, channel and source coding, digital modulation techniques, and development of handicap aids for children with speech problems.