A STUDY OF THREE CODERS (SUB-BAND, RELP AND MPE) FOR SPEECH WITH ADDITIVE WHITE NOISE

K.K. PALIWAL* AND T. SVENDSEN

DIVISION OF TELECOMMUNICATIONS
UNIVERSITY OF TRONDHEIM, TRONDHEIM-NTH, NORWAY

(*PRESENT ADDRESS: COMPUTER SYSTEMS AND COMMUNICATIONS GROUP, TATA INSTITUTE OF FUNDAMENTAL RESEARCH, HOMI BHABHA ROAD, BOMBAY-400 005, INDIA)

ABSTRACT.

The following three speech coders are implemented for a bitrate of 9.6 kbits/s 1) Sub-band coder, 2) Residual Excited Linear Predictive (RELP) coder, and 3) Multi-Pulse Excited linear predictive (MPE) coder. Performance of these coders is evaluated for speech corrupted by additive white noise. Evaluation of speech coders is done both subjectively and objectively. The MPE coder is found to give the best performance among the three coders. It is also shown that the MPE coder can be used for noisy speech with signal-to-noise ratio as low as -10 dB giving reasonably good quality speech provided 1) one does not use the error weighting filter and 2) one can use a better LP analysis algorithm which can estimate LP coefficients correctly from noisy speech.

1. INTRODUCTION

Recent advances in digital modulation techniques have made it feasible to transmit speech in digital form for mobile radio telephony. (See reference [1] for some of these digital modulation techniques). With the limited bandwidth of about 25 kHz available for the mobile telephone systems, these new digital modulation techniques put an upper bound on the speech transmission rate which is about 16 kbits/s. Several speech coding techniques which satisfy this requrement have been reported in the literature [2]. These medium bitrate speech coders (9-16 kbits/s) yield reasonably good quality speech when operating in quiet environments. However, in practice the environments in which these coders have to operate are noisy; for example, the cockpit of an aeroplane or inside a moving car.

The aim of the present paper is to study these speech coders for noisy speech. As a first approximation to environmental noise we study speech corrupted by additive white noise.

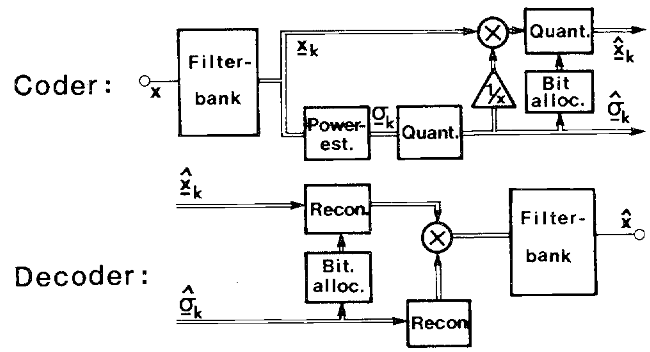

Fig.1 Sub-band coder

The following three promising medium bitrate speech coders are selected for investigation in the present study: 1) Sub-band coder [3], 2) Residual Excited Linear Predictive (RELP) coder [4] and 3) Multi-Pulse Excited linear predictive (MPE) coder [5]. These coders were implemented at a bitrate of 9.6 kbits/s. Noisy speech is obtained by first generating white Gaussian noise and then adding it to speech recorded in a quiet environment. Peformance of speech coders is evaluated in terms of objective signal-to-noise ratio (SNR) and through informal listening tests.

2. SPEECH CODERS

In this section, we describe the three speech coders (Sub-band, RELP and MPE) implemented for a bitrate of 9.6 kbits/s. These coders are simulated in FORTRAN on a HP-1000 minicomputer using speech digitized at 8 kHz sampling rate as input signal. Block diagrams of the coders are shown in figures 1,2 and 3.

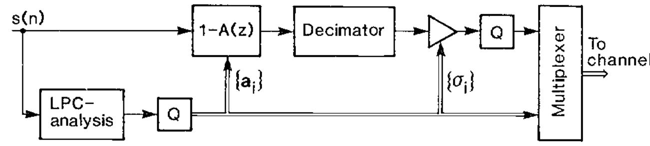

Fig.2 RELP coder

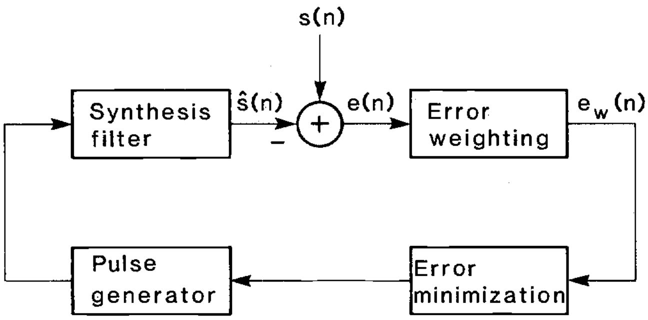

Fig.3 MPE coder

Sub-band coder: The input signal is divided into 14 sub-bands by the use of a filterbank consisting of quadrature mirror filters. Each band has a bandwidth of 250Hz and the frequency range covered is 0-3500 Hz. The sub-band signals are then organized in blocks of 16 ms duration. For every block an estimate of the power in each band is made. These estimates are quantized and used for power normalization and for determining the number of bits to be allocated to each band [3]. The power estimates are transmitted to the receiver as side information. A Max-Lloyd quantizer is then employed to perform the quantization of the sub-band signals

RELP coder: The RELP coder organizes the input signal in blocks of 27 ms duration. For each block a 10th order LPC analysis is performed. The quantized LPC coefficients are used for whitening the signal spectrum. After whitening, the signal is low-pass filtered with a cut-off frequency of 1 kHz and decimated by a factor of 4. Before quantizing, the base-band signal is divided into 3 sub-blocks of equal length. The power of each of the sub-blocks is estimated and used for power normalization and as input to a simple bit allocation algorithm which determines the number of bits to be assigned to each sub-block [4]. At the receiver end the high-frequency portion of the signal is regenerated by the perturbed spectral folding method and the reconstructed speech is generated by passing the fullband residual signal through the LPC synthesis filter.

MPE coder: The input signals is divided into blocks of 10 ms duration. For each block a 10th order LPC analysis is performed, using an analysis frame of 20 ms. The LPC analysis gives the coefficients of the prediction filter, $H(z)=1-A(z)$. The error weighting filter used in the optimization is of the form $W(z)=(1-A(z))/(1-A(z/\gamma))$ with $\gamma=0.8$ as the standard value. The basic version of the coder operates similarly to the MPE-coder described in [5], using a minimum mean squared weighted error criterion for finding the optimum pulse positions and amplitudes. For each block of speech, 8 pulses are found. The positions are quantized using the combinatorial method of [6]. The amplitudes are normalized by the greatest amplitude found before quantizing.

3. PERFORMANCE EVALUATION AND RESULTS

For evaluating the performance of the three speech coders (Sub- band, RELP and MPE), two different English sentences are used; one spoken by a male speaker and the other by a female speaker. Recording of these sentences was done in a quiet room. The speech signal was digitized at a sampling frequency of 8 kHz by a 16-bit analog-to-digital converter.

Noisy speech which is used as input to the speech coder is obtained by first generating zero-mean white Gaussian noise and then adding it to the original (noise-free) speech. The signal-to-noise ratio (SNR) of noisy speech is defined as

$$SNR_{in} = 10 \log_{10} \frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N} [x(n) - y(n)]^2} \qquad (1)$$

where x(n) and y(n) are the n'th samples of the original and noisy speech, respectively, and N is the total number of samples in the speech signal.

The speech coders accept the noisy speech $y(n)$, $n=1,\ldots,N$, as input. It encodes it at the transmitting end and decodes (or reconstructs) it at the receiving end. In order to evaluate the performance of the speech coders objectively, we use the SNR of the reconstructed speech, which is defined as follows:

$$SNR_{out} = 10 \log_{10} \frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N} [x(n) - z(n)]^2} \qquad (2)$$

where $z(n)$, $n=1,\ldots,N$, is the reconstructed speech signal.It must be emphasized that the SNR values do not always give a correct picture of the subjective quality. Because of this, we also use informal listening of reconstructed speech to make subjective evaluation of the performance of the speech coders.

We have studied the performance of the three speech coders for noisy input speech at the following SNR values: ∞dB

Table I

Effect of additive white noise on the performance of the MPE coder:

Input speech SNR (in dB)	SNR (in dB) of reconstructed speech		
	Sub-band Coder	RELP Coder	MPE Coder
∞	12.12	6.39	10.75
30	12.08	6.74	10.83
20	11.80	6.77	10.29
10	9.05	5.74	7.88
0	1.70	0.32	1.79
- 10	- 7.54	- 8.81	- 6.82

(corresponding to original noise-free speech), 30dB, 20dB, 10dB, 0dB and -10dB. The SNR values of the reconstructed speech signal resulting from the Sub-band, RELP and MPE coder are listed in Table I. The following observations are made from this table and through the informal listening test. For input SNRs above 15 dB, although the Sub-band coder yields the highest SNR, subjectively the MPE coder yields the highest speech quality and the RELP coder the lowest. For input SNRs below 15 dB, the performance of all the three coders degrades drastically. At these low SNRs, differences in speech quality resulting from the three coders are very small, although the RELP coder sounds inferior to the other two coders.

Thus, from this evaluation using noisy speech, we can conclude that the MPE coder performs slightly better than the other two coders. This motivated us to study the MPE coder in more detail and to see whether we can improve its performance for noisy speech.

Table II

Effect of using simultaneous optimization of pulse amplitudes on the performance of the MPE coder.

Input speech SNR (in dB)	SNR (in dB) of reconstructed speech	
	MPE coder	MPE coder with simultaneous optimization of amplitudes
∞	10.75	11.28
30	10.83	11.07
20	10.29	10.62
10	7.88	7.83
0	1.79	1.86
- 10	- 6.82	- 6.81

Recently a modified version of the MPE coder has been proposed by Singhal and Atal [6] and Berouti et al [7]. In this modified version, the pulse amplitudes are simultaneously optimized each time a new pulse location is found in the given speech record. We have incorporated this modification in our MPE coder. The modified version was studied for different SNR values of input noisy speech and the results are listed in Table II. We observe that the modified version slightly improves the speech quality of MPE coder for input SNRs greater than 20dB. But, for input SNRs less then 20dB no improvement is noticable. Also, the modified version is computationally very expensive. So we can say that the modified MPE coder does not help much in solving the problem for noisy speech.

In the MPE coder, a weighting filter (with weighting factor $\gamma=0.8$) is used to shape the residual error signal for improving the subjective quality of speech [5]. In the present study, we also studied the MPE coder without this error weighting filter (i.e. $\gamma=1.0$).

The SNR values of the reconstructed speech under different noise conditions are listed in Table III. We observe from the table and through the subjective listening that though the error weighting filter helps in giving better quality speech (subjectively) at input SNRs above 20dB, it degrades the performance of the MPE coder (both subjectively and objectively) for noisy speech with SNRs below 20dB.

Table III

Effect of error weighting filter on the performance of the MPE coder.

Input speech SNR (in dB)	SNR (in dB) of reconstructed speech	
	MPE coder with error weighting filter ($\gamma = 0.8$)	MPE coder without error weighting filter ($\gamma = 1.0$)
∞	10.75	11.14
30	10.83	11.12
20	10.29	10.64
10	7.88	8.42
0	1.79	1.88
- 10	- 6.82	- 7.02

In order to see whether one can improve the performance of the MPE coder for noisy speech by employing a better algorithm for estimating the LP coefficients, we stored the LP coefficients computed from the original noise-free speech and used them for noisy speech to find the pulse

Table IV

Performance of the MPE coder when correct LP coefficients derived from original (noise-free) speech are used.

Input speech SNR (in dB)	SNR (in dB) of reconstructed speech	
	MPE coder with error weighting filter ($\gamma = 0.8$)	MPE coder without error weighting filter ($\gamma = 1.0$)
∞	10.75	11.14
30	10.74	11.28
20	10.41	11.17
10	7.17	10.29
0	- 0.66	5.49
- 10	- 9.96	- 3.29

locations and amplitudes at the transmitting end and to synthesize speech at the receiving end. The objective results for different input SNRs are listed in Table IV. By comparing the objective results listed in this table with those listed in Table III and by conducting the informal listening test, we observe that this procedure of employing LP coefficients computed from original speech gives no improvement in speech quality (both subjectively and objectively) for the MPE coder using the error weighting filter (i.e. γ=0.8). However, when the MPE coder was used without the weighting filter (i.e. γ=1.0), this procedure significantly improved the speech quality (both subjectively and objectively). Note, however, that γ=1.0 will optimize the SNR whereas γ<1 uses a weighted error criterion and thus yields lower SNRs. Thus we can conclude that the MPE coder can be used for noisy speech with SNR as low as -10dB giving reasonably good quality speech provided we do not use the error weighting filter and we can employ an LP analysis method which gives correct estimates of LP coefficients for noisy speech.

Recently a number of LP analysis methods have been reported in the literature [8-10] which can give better estimates of LP coefficients for noisy speech. We also have developed a new algorithm of LP analysis for noisy speech using the extended Yule-Walker equations [11]. We have used it with MPE coder and the preliminary results are found to be encouraging. The detailed results will be reported elsewhere in future.

6. CONCLUSION

The three medium bitrate speech coders (Sub-band, RELP and MPE) implemented at 9.6 kbits/s were studied as to their performance for speech degraded by additive white noise. Both subjective and objective tests were used to evaluate the performance of these coders. The MPE coder was found to give the best performance among the three coders. The MPE coder was studied in more detail and it was shown that we can use this coder for noisy speech with SNR as low as -10dB and can still get reasonably good quality speech provided 1) we do not use the error weighting filter and 2) we can find an LP analysis algorithm which can estimate the LP coefficients correctly from noisy speech.

REFERENCES

[1] IEEE Trans Commun., Special Section on Combined Modulation and Encoding, Vol. COM-29, Mar. 1981.

[2] N.S. Jayant and P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Prentice- Hall, New Jersey, 1984.

[3] T.A. Ramstad, "Sub-band coder with a simple adaptive bit-allocation: A possible candidate for digital mobile telephony?", Proc. ICASSP, 1982, pp. 203-207.

[4] T. Svendsen, "Effects of Rayleigh fading on medium bandwith speech coders", Proc. EUSIPCO-83, 1983, pp. 539-542.

[5] B.S. Atal and J.T. Remde, "A new model of LPC excitation for producing natural sounding speech at low bit rates". Proc. ICASSP, 1982, pp. 614-617.

[6] S. Singhal and B.S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates", Proc. ICASSP, 1984, pp. 1.3.1 - 1.3.4.

[7] M. Berouti et al., "Efficient computation and encoding of the multipulse excitation for LPC", Proc. ICASSP, 1984, pp. 10.1.1 - 10.1.6.

[8] J.A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach", Proc. IEEE, Vol.70, No. 9, Sept. 1982, pp. 907-939.

[9] J.S. Lim and A.V. Oppenheim, "All-pole modeling of degraded speech", IEEE Trans. Acoust. Speech and Signal Processing, Vol. ASSP-26, No.3, June 1978, pp. 197-210.

[10] R.W. Johnson and J.E. Shore, "Minimum cross-entropy spectral analysis of multiple signals", IEEE Trans. Acoust. Speech and Signal Processing, Vol. ASSP-31, No. 3, June 1983, pp. 574-582.

[11] K.K. Paliwal, "Two new AR spectrum estimation algorithms using extended Yule-Walker equations for noisy signals", to be published.