FAST STOCHASTIC CODEBOOK SEARCH THROUGH THE USE OF ODD-SYMMETRIC CROSSCORRELATION BASIS VECTORS

Cheung-Fat Chan
Dept. of Electronic Engineering
City University of Hong Kong
83 Tat Chee Avenue, Kowloon, HONG KONG
E-Mail: eecfchan@cityu.edu.hk

ABSTRACT

A fast codebook search method for code excited linear predictive (CELP) coding of speech is described. The method relies on using a vector-sum codebook where the crosscorrelation of any pair of basis vectors is odd-symmetric. Due to this odd-symmetric crosscorrelation (OSC) property the energy term of the cost function for codebook search is a constant with respect to the search, and a simple sign detection procedure is used to locate the optimum codeword with a complexity almost independent of the codebook size. An algorithm for generating a set of OSC basis vectors will be described. Simulation results show that by replacing the standard VSELP codebooks with the OSC codebooks, same coder performance can be achieved but with a much lower complexity. An OSC-CELP coder was implemented and demonstrated to achieve good quality speech at rates below 4.8 kbps.

1. INTRODUCTION

In CELP coding the "best" codevector is chosen from the excitation codebook by maximizing the cost function [1],

$$\frac{(u^T H^T x)^2}{u^T H^T H u} \tag{1}$$

where u is the codevector from the excitation codebook, H is the impulse response matrix of the weighted LPC synthesis filter, and x is the weighted input speech after the removal of zero-state response. The complexity in searching a codebook using the cost function of (1) is dominated by the calculation of the denominator, because, in each evaluation of the cost function, a filtering process has to be performed. After many years of intensive research in CELP coding, several fast algorithms were developed; noticeable of these are algebraic CELP[2] and VSELP[3]. In vector-sum codebook, the codeword u is generated as a linear combination of a set of basis vectors, $u = [w_1 \ w_2 \ . \ . \ . \ w_M] b = Wb$ where w_1 to w_M are the basis vectors of dimension N $(M<N)$, b is an M-dimension column vector with elements being either 1 or -1. By using the vector-sum codebook, the cost function of (1) becomes

$$\frac{(b^T v)^2}{b^T A b} \tag{2}$$

where $v = W^T H^T x$ and $A = W^T H^T H W$. It has been shown that, by using vector-sum codebooks together with Gray code sequencing of the binary vector b, codebook search can be performed with a complexity of $O(M2^M)$ instead of $O(N^2 2^M)$ normally required by conventional method[3]. However, even using Gray code search, the complexity is still linearly proportional to the codebook size, i.e., 2^M. It is obvious that if A is a diagonal matrix with diagonal elements a_{ii}, then, the term in the denominator of (2) will be a constant irrespective of the binary vector b, i.e.,

$$b^T A b = \sum_{i=1}^{M} (b_i)^2 a_{ii} = \sum_{i=1}^{M} a_{ii} \tag{3}$$

with b_i taking a value of 1 or -1. If $b^T A b$ is a constant with respect to the search, then the cost function of (2) is maximized iff $(b^T v)^2$ is maximized. This is simply achieved by detecting the sign of the element v_i in v, i.e., $b_i = sgn(v_i)$ for $1 \le i \le M$, where $sgn(*) = 1$ if the argument * is positive or $sgn(*) = -1$ is the argument * is negative. The complexity of searching this codebook is $O(M)$ and is almost independent of the codebook size. Some methods have been proposed in the past to approximate A as a diagonal matrix, for examples, sparse binary pulse excitation[4], sparse algebraic codes[2]. These methods, however, depend on the sparsity of pulses, and this in terms would directly affect the synthetic speech quality if the sparsity of pulses is high. In this paper, we propose a method to design a set of basis vectors such that A is always diagonal.

2. Basic Vectors With Odd-Symmetric Crosscorrelation

It has been shown in [1] that the covariance matrix $C = H^T H$ can be approximated by an autocorrelation matrix which is symmetric Toeplitz if the impulse response of the weighted synthesis filter is properly truncated. Then using the fact that C is a Toeplitz matrix, the element a_{ij} in A can be calculated as

$$a_{ij} = w_i^T C w_j = c(0) r_{ij}(0) + \sum_{n=1}^{N-1} c(n) \left[r_{ij}(n) + r_{ji}(n) \right] \tag{4}$$

where $c(n)$ is the autocorrelation of the impulse response of the weighted synthesis filter, and $r_{ij}(n) = w_i^T Z^n w_j$ is the crosscorrelation of basis vectors w_i and w_j. Note that Z is a shift (delay) matrix. It is straight forward to show that $r_{ji}(n) = r_{ij}(-n)$. In order to make A diagonal, we must have $a_{ij} = 0$ for $i \neq j$. Obviously, from (4), this is achieved if $r_{ij}(0) = 0$ and $r_{ij}(n) + r_{ij}(-n) = 0$ for $1 \le n \le N-1$, $1 \le i,j \le M$ and $i \neq j$ In other words, if we can design a set of M basis vectors such that the crosscorrelation of any pair of vectors is odd-symmetric, then A will be diagonal irrespective of C. In order to satisfy the OSC criteria for any pair of vectors, the total number of simultaneous equations needed to be related is $NM(M-1)/2$. If we add an additional constraint that the energy of any basis vector is unity, i.e., $\|w_i\|^2 = 1$ for $1 \le i \le M$, then, the total number of equations involved becomes $M + NM(M-1)/2$. But, there are totally MN components (N elements in each vector) in these equations. By using the fact that these equations must not be under-determined for the existence of OSC vector sets, we immediately conclude that the inequality $M \le (3N-2)/N$ must be satisfied in order to obtain OSC basis vectors. From this result we see that $M<3$ for $N>0$. Obviously, we can design an OSC codebook with a maximum of only *two* basis vectors! For practical CELP coders, the number of basis vectors required is normally between 7 to 12. Fortunately, we can relax the OSC constraint by noting that the autocorrelation sequence $c(n)$ will approach zero for a sufficient large n, say L. It is not necessarily to design a set of vectors such that the OSC constraint is enforced all the way up to length N-1. By using this assumption, a modified criteria is derived and we have the following inequality

$$L \le \frac{2N-2}{M-1} - 1 \qquad (5)$$

For a typical value of $N=40$ and $M=10$, we have $L \le 7$. Although $L=7$ seems to be small, C is heavily banded, and it is experimentally determined that, with a perceptual weighing factor of 0.8, $c(n)$ will drop to less than 10 percent of $c(0)$ when $n>7$. Fig. 1 shows the ensemble averages of $|c(n)|$ against n for various weighing factors γ. In this test, the weighting filter is defined as $A(z)/A(z/\gamma)$ where $A(z)$ is the LPC inverse filter. Therefore, provide that we can design a codebook with (shorten) OSC property, it is still viable to assume A to be diagonal and to utilize the fast procedure mentioned previously to determine the codeword.

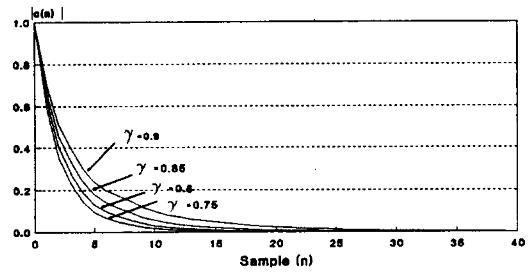

Fig. 1 Ensemble Averages of $|c(n)|$

3. Generation of the OSC Basis Vectors

Theoretically, it is quite trivial to find a set of vectors that mutually satisfy the OSC criteria. In this work, a procedure is utilized to optimize a set of random vectors such that their crosscorrelations become mutually odd-symmetric. The procedure is based on iteratively minimizing the error function ξ_{osc} defined in (6) by adjusting the vectors w_i for $1 \le i \le M$ sequentially according to their gradient directions.

$$\begin{aligned}\xi_{osc} &= \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\ne i}}^{M}\left\{r_{ij}^2(0) + \sum_{n=1}^{L}\left[r_{ij}(n)+r_{ij}(-n)\right]^2\right\}\\ &= \sum_{i=1}^{M}\sum_{\substack{j=1\\ j\ne i}}^{M}\left\{w_i^T w_j w_j^T w_i + \sum_{n=1}^{L} w_i^T[Z^n+(Z^n)^T]w_j w_j^T[Z^n+(Z^n)^T]w_i\right\} \qquad (6)\\ &= \sum_{i=1}^{M} w_i^T\left[\sum_{\substack{j=1\\ j\ne i}}^{M}\Phi_j\right]w_i\end{aligned}$$

$$\text{where } \Phi_j = w_j w_j^T + \sum_{n=1}^{L}[Z^n+(Z^n)^T]w_j w_j^T[Z^n+(Z^n)^T] \qquad (7)$$

is a symmetric matrix. Note that L is the length of the crosscorrelation sequence that will satisfy the odd-symmetry property. Using the fact that $w_j^T\Phi_i w_j = w_i^T\Phi_j w_i$ the gradient due to vector w_i is thus calculated as

$$\frac{\partial \xi_{osc}}{\partial w_i} = 4\left[\sum_{\substack{j=1\\ j\ne i}}^{M}\Phi_j\right]w_i \qquad (8)$$

The algorithm for generating OSC vectors is given in A1

Initialize w_i *for* $1 \le i \le M$ *as random vectors*

$$\begin{aligned}&R = 4\sum_{j=2}^{M}\Phi_j\\ &\textit{Repeat}\\ &\quad \textit{For } i=1 \textit{ to } M \textit{ do}\\ &\qquad w_i = [I - \lambda R]w_i\\ &\qquad w_i = \frac{w_i}{\|w_i\|}\\ &\qquad R = R + 4\Phi_i - 4\Phi_{(i)_M+1}\\ &\textit{Until } \xi_{osc} = 0\end{aligned} \qquad (A1)$$

where Φ_i and ξ_{osc} are defined in (7) and (6), respectively, and λ is the adaptation step size which controls the speed of convergence. It is also worth while mentioning that the self-orthogonal (SO) criteria [5] can also be imposed on the basis vectors. For self-orthogonal vectors, the criteria of $r_{ii}(n) = 0$ for $1 \le n \le N-1$ must be satisfied. The advantage of imposing the SO criteria is that matrix A will become an identity matrix as $w_i^T C w_i = c(0) r_{ii}(0) = 1$ if $c(0)=1$ and $r_{ii}(0) = \|w_i\|^2 = 1$. Then, for $A = I$, it is not necessary to calculate the denominator term in (2) for the determination of the excitation gain after the codebook search since $b^T b = M$. With an additional constraint on self-orthogonality, extra ML equations are needed to be considered. Then, a new inequality for the condition on the existence of SO-plus-OSC vectors is derived as

$$L \le \frac{2N}{M+1} - 1 \qquad (9)$$

Again, for a typical value of $N=40$ and $M=10$, we have $L \le 6$. The additional error due to the self-orthogonal constraint is calculated as

$$\xi_{so} = \sum_{i=1}^{M}\sum_{n=1}^{L} w_i^T Z^n w_i w_i^T (Z^n)^T w_i \qquad (10)$$

The corresponding offset of the gradient vector due to w_i is $\partial\xi_{so}/\partial w_i = 2\Delta_i w_i$, where

$$\Delta_i = \sum_{n=1}^{L}(w_i^T Z^n w_i)[Z^n + (Z^n)^T] \qquad (11)$$

The algorithm for generating SO-plus-OSC vectors is given as follows:

Initialize w_i *for* $1 \le i \le M$ *as random vectors*

$$\begin{aligned}&R = 4\sum_{j=2}^{M}\Phi_j\\ &G = R + 2\Delta_1\\ &\textit{Repeat}\\ &\quad \textit{For } i=1 \textit{ to } M \textit{ do}\\ &\qquad w_i = [I - \lambda G]w_i\\ &\qquad w_i = \frac{w_i}{\|w_i\|}\\ &\qquad R = R + 4\Phi_i - 4\Phi_{(i)_M+1}\\ &\qquad G = R + 2\Delta_i\\ &\textit{Until } (\xi_{osc} + \xi_{so}) = 0\end{aligned} \qquad (A2)$$

where Δ_i and ξ_{so} are defined in (11) and (10), respectively.

4. Properties of the OSC Basis Vectors

Fig. 2 shows the time waveforms and spectra of the OSC basis vectors. These vectors were obtained by using the OSC vectors generation algorithm (A1) with the setting; $N=40$, $M=8$, and

$L=7$. It can be seen from the spectrum plots that, although the spectra of OSC vectors are not white, they contain a wide variety of spectral components. Therefore, it is expected that OSC vectors are well suitable for use as basis vectors for constructing stochastic excitation signals.

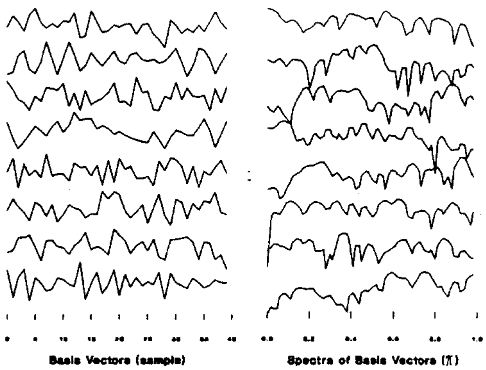

Fig. 2 OSC Basis Vectors and Their Spectra (N=40, M=8, L=7)

Fig. 3 shows the plots of $r_{ij}(n)$ and $r_{ij}(n)+r_{ji}(n)$ for the basis vectors shown in Fig. 2. Due to the odd-symmetric property, it can be seen from the plots in Fig. 3 that $r_{ij}(n)+r_{ji}(n)=0$ for $|n|\leq 7$ (between 2 vertical lines).

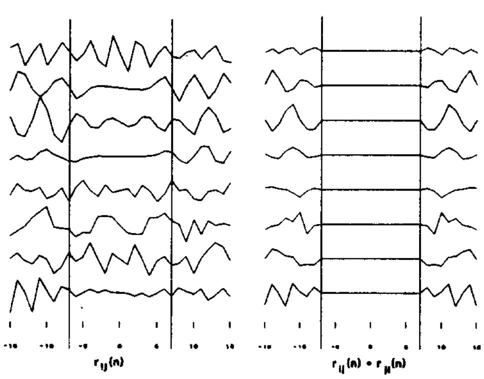

Fig. 3 Plots of $r_{ij}(n)$ and $r_{ij}(n)$ + $r_{ji}(n)$ for Basis Vectors Shown in Fig. 2 (only 8 out of 28 crosscorrelations are shown)

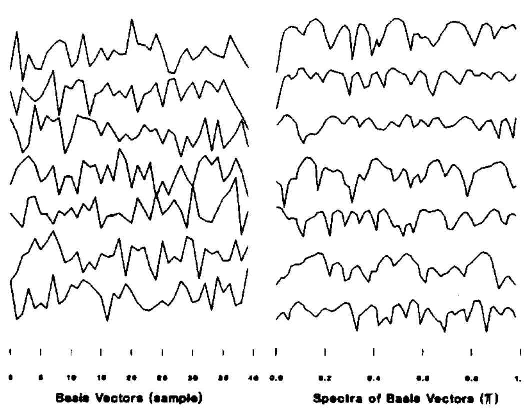

Fig. 4 SO-plus-OSC Basis Vectors and Their Spectra (N=40, M=7, L=5)

Fig. 4 shows the time waveforms and spectra of the SO-plus-OSC vectors generated by using Algorithm (A2) with the setting: $N=40$, $M=7$ and $L=5$. From these plots we see that the spectrum of SO-plus-OSC vector is white. This agrees well with the theory because the autocorrelation of white noise sequence is a delta function. It should be reasonably to assume that SO-plus-OSC vectors are more suitable than OSC vectors for constructing stochastic excitation signals. Several observations about the properties of the OSC vectors can be made. First, since $r_{ij}(0)=0$ for $i\neq j$, OSC basis vectors are also mutually orthogonal (i.e. $W^TW=I$). This orthogonal property aligns well with other methods which use transformed excitation vectors[2,4].

5. Simulation Results

The simulation conditions are listed as follows: speech signal is bandlimited and sampled at 8 kHz using a 16-bit A/D converter. A classic CELP coder structure with an adaptive codebook (7-bit integer pitch) and a stochastic codebook (128 codewords) is used in this test. Short-term linear predictive analysis is performed using autocorrelation method with 32ms Hamming window and 20ms frame size. The perceptual weighting factor is set to 0.8. Codebook searches are performed in closed-loop with adaptive codebook search taking the precedent. The excitation parameters are determined in every subframe of 5ms (40 samples) long. Four OSC codebooks are used in the test and they were generated using Algorithm (A1) with different length of constraint on odd-symmetry crosscorrelation. Table 1 lists the percentage occurrence of Hamming Distance (HD) of codewords obtained by using the fast procedure and the exhaustive search. It is clear from these results that the occurrence of HD is directly related to L. For $L=8$, over 88% of codewords found using the fast procedure are the same as those found by exhaustive search, and only 11% of codewords have 1 bit in difference.

M=7,N=40	HD=0	HD=1	HD=2	HD=3	HD=4	HD=5	HD=6
L=8	88.2	11.2	0.6	0	0	0	0
L=7	87.5	11.7	0.7	0.1	0	0	0
L=6	83.6	13.9	2.3	0.2	0	0	0
L=5	82.7	15.7	1.5	0.1	0	0	0

Table 1 Percentage Occurrence of Hamming Distances of Codewords Obtained Using Fast and Exhaustive Search

In order to compare the performance of coders using OSC codebooks with conventional CELP coder using Gaussian codebooks, we implemented several coders based on the 7.95kbps VSELP standard structure defined in [3]. In this structure, two stochastic codebooks of 7 bits each are used. The codevector dimension is 40. The excitation gains are vector quantized to 8 bits. In this experiment, totally 5 coders were implemented. Coder 1 is the standard VSELP coder which employs two Gaussian codebooks[3]. Coder 2 and coder 3 have the same structure as coder 1 but the Gaussian codebooks are replaced by two OSC codebooks which were independently generated with the OSC constraint length of $L=7$, however, coder 2 utilizes the fast procedure based on diagonal approximation of A while coder 3 utilizes an exhaustive search. Coder 4 and 5 are the same as coder 2 and 3 excepted that SO-plus-OSC codebooks generated with $L=5$ were used. Table 2 shows the SNR performance of these coders. The SegSNR results indicate that OSC codebooks achieve the same performance as for Gaussian codebooks. The drop in SegSNR due to the use of fast procedure based on diagonal approximation is very small; only 0.07 dB in case of OSC codebooks and 0.09 dB in case of SO-plus-OSC codebooks. This result also reveals that, even with a shorter OSC constraint length, SO-plus-OSC codebooks outperform OSC codebooks and Gaussian codebooks.

	Conventional VSELP (Gaussian Codebook)	OSC Codebook Fast Search (L=7)	OSC Codebook Exhaustive Search (L=7)	SO+OSC Codebook Fast Search (L=5)	SO+OSC Codebook Exhaustive Search (L=5)
SNR	13.04	13.05	13.15	13.27	13.62
SegSNR	13.85	13.60	13.67	14.22	14.31

Table 2 Performance Comparisons of OSC Codebooks and Gaussian Codebooks

6. OSC-CELP Coder at Rates Below 4.8 kbps

Fig. 5 shows the block diagram of the proposed OSC-CELP coder. The short-term LP analysis is similar to other analysis-by-synthesis coders. The excitation is constructed from a selectable combination of adaptive and stochastic contributions. There are 4 possible combinations and the one that achieves the minimum (closed-loop) distortion is selected during encoding.

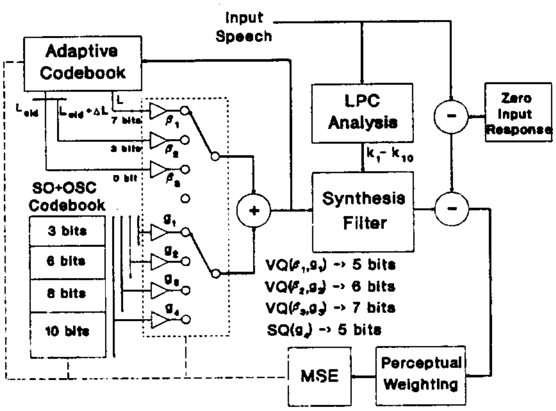

Fig. 5 Block Diagram of the Proposed OSC-CELP Coder

The rationale behind the proposal of the selectable combination of excitation source is to obtain a flexible way to allocate scare resource to code the excitation parameters. Because of very low complexity in searching the OSC codebook, the scheme retains the robustness of closed-loop minimization without relying on using an open-loop classifier. The stochastic codebook is a 10-bits SO-plus-OSC codebook generated using Algorithm (A2). The stochastic codebooks for Modes 1 to 3 are the subsets of the 10-bits SO-plus-OSC codebook. Codebook search is performed using the fast procedure based on the diagonal approximation and a HD-1 correction is applied after the search. The short-term LP analysis (order 10) is performed using the autocorrelation method with Hamming window. The PARCOR coefficients are calculated and quantized to 35 bits using a trellis in-loop quantization scheme[6]. There are 5 subframes in each LP frame. An OSC-CELP coder which can operate at rates of 4.8 kbps, 4 kbps, and 3.2 kbps was implemented. The coder has different subframe sizes at different rates. At 4.8 kbps, the coder has a subframe size of 40 samples and the OSC codebook was generated using a SO-plus-OSC constraint length of $L=4$. At 4 kbps, the coder has a subframe size of 48 samples and the OSC codebook was generated using a SO-plus-OSC constraint length of $L=5$. At 3.2 kbps, the coder has a subframe size of 60 samples and the OSC codebook was generated using a SO-plus-OSC constraint length of $L=6$. The bit allocation is listed in Table 3.

Spectrum (k_1-k_{10})	35					35
Mode Selection	2 x 5					10
Excitation	Gain		Shape			
Mode 1	5	+	7+3	=	15 x 5	75
Mode 2	6	+	3+6	=	15 x 5	75
Mode 3	7	+	0+8	=	15 x 5	75
Mode 4	5	+	10	=	15 x 5	75

Table 3 Bit Allocation of the Proposed OSC-CELP Coder (Total 120 bits per frame)

Table 4 lists the SNR performance of the coder at different rates. The proposed OSC-CELP coder achieves a SegSNR of 12.61 dB at 4.8 kbps which is better than the 4.8 kbps DoD FS1016 CELP coder which scores 11.33 dB. We found that the speech quality at 4 kbps is almost the same as the 4.8 kbps DoD CELP. At 3.2 kbps, the coder produces speech which sounds a little bit rough but is considered good quality with high intelligibility. The OSC-CELP coder has been successfully implemented in real-time on Motorola DSP56156 16-bit fixed-point DSP (40 MHz with zero wait state) which occupies 2.5 k of program ROM and 1.8 k of data RAM. The coder achieves full-duplex operation with only 50% processor loading (35% encoding and 15% decoding).

	4.8 kbps	4 kbps	3.2 kbps
SNR	12.12	10.79	9.87
SegSNR	12.61	11.36	10.15

Table 4 SNR Performance of the Proposed OSC-CELP Coder

7. Conclusion

This paper presents a method to design vector-sum codebook where the crosscorrelation of any pair of basis vectors in the codebook is odd-symmetric. As a result of odd-symmetry property, codebook search can be performed in a very efficient manner. Simulation results indicate that VSELP coder which employs the OSC codebooks can achieve comparable performance with standard VSELP coder which employs Gaussian codebooks. An OSC-CELP coder was successfully implemented and demonstrated to achieve good-quality reproduction speech at bit rates below 4.8 kbps.

References

1. I.M. Trancoso and B. S. Atal, "Efficient Search Procedures for Selecting the Optimum Innovation in Stochastic Coders," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 38, No. 3, pp.385-396, March 1990.
2. J. P. Adoul, P. Mabilleau, M. Delprat, and S. Morisette, "Fast CELP Coding Based on Algebraic Codes," Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing, pp.1957-1960, 1987.
3. I A. Gerson and M.A. Jasiuk, "Vector Sum Excited Linear Predictive (VSELP) Speech Coding at 8 kbps," Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing, pp.461-464, 1990.
4. R. Salami and D. Appelby, "A New Approach to Low Bit Rate Speech Coding with Low Complexity using Binary Pulse Excitation (BPE)," IEEE Workshop on Speech Coding for Telecommunications 1989.
5. K.W. Law, W.F. Leung and C.F. Chan, "On Reducing the Complexity and Storage of CELP Speech Coding Using Self-Orthogonal Codebook," IEEE Workshop on Speech Coding for Telecommunications, pp.103-104, 1993.
6. K.W. Law and C.F. Chan, "Trellis In-Loop PARCOR Coefficients Quantizer for Speech Signals," IEE Proceedings in Communications, Speech, and Vision, Vol.140, No. 5, pp.321-326, Oct. 1993.