ON MSE OF CELP CODER

Jinho Choi

Speech Team, Multimedia Lab.
DACOM Corporation / R&D Center
34 Gajeong Dong, Yuseung Gu, Daejeon, 305-350, Korea
Tel: +82-42-220-4225, Fax: +82-42-220-4277, 4444
email: jin@halla.dacom.co.kr

ABSTRACT - We consider mean squared error (MSE) between original and synthesized speech signals in terms of length of frame and size of stochastic codebook of CELP coder. A theoretical approximation of MSE has been found in terms of length of frame and size of stochastic codebook. It is shown that MSE depends on not only the length of frame, but also correlation properties of speech signal when the size of stochastic codebook is fixed. Moreover, relationship between MSE and correlation properties of speech signal is observed using the number of effective dimensions of speech signal. For example, uncorrelated speech signal which has large number of effective dimensions has larger MSE than those of correlated speech signal which has small number of effective dimensions. In addition, two applications of MSE are considered to reduce stochastic codebook search time. One application can be also used for reduction of number of bits to encode codeword of stochastic codebook.

INTRODUCTION

Speech signal has been compressed by various coders, e.g., adaptive differential pulse code modulation (ADPCM), linear predictive coder (LPC), and, recently, code excitation linear prediction (CELP) coder [4]. The CELP coder can be characterized by utilization of excitation signal. In general, two excitation signals have been used [1]. One is self-excitation signal and the other is stochastic excitation signal. Each excitation signal, codeword, can be generated from codebook. For self-excitation and stochastic excitation codewords, adaptive self-excitation and stochastic codebooks are used, respectively. The quality of synthesized speech can be enhanced by increasing the size of each codebook. Since the computational complexity of encoding heavily depends on the size of codebook, we cannot use large size of codebooks, i.e., there is trade off between the quality of synthesized speech and computational complexity. In [3][6], fast methods to determine codeword from codebook have been proposed without decreasing quality of synthesized speech. Using structured codebook, another CELP coder, vector sum excited linear predictive (VSELP) coder [5], has been established and it is shown that VSELP coder has small computational complexity to determine codeword. In addition, to minimize coding delay of CELP coder, low delay CELP (LD-CELP) coder has been developed in [2]. Since computational complexity and coding delay of CELP and LD-CELP coders depend on the size of stochastic codebook, the determination of size of stochastic codebook is of importance.

Suppose that speech signal can be modeled by autoregressive (AR) model. For AR(Q) model, signal $y(n)$ has the relationship which is given by

$$y(n) + \sum_{k=1}^{Q} a_k y(n-k) = e(n), \tag{1}$$

where $\{a_k\}$ are AR coefficients and $e(n)$ is independent identically distributed (iid) random sequence with mean zero and variance σ^2.

We will now formulate analysis-by-synthesis procedure using matrix notation. By ignoring initial values and error dues to finite impulse response (FIR) approximation, the filter output of the mth codeword, $\mathbf{y}_m \in R^{N\times 1}$ (where $R^{N\times 1}$ is the space $N \times 1$ real valued column vectors) with its nth component given by $y_m(n)$, in the present frame is given by

$$\mathbf{y}_m = \alpha_m H \mathbf{e}_m, \qquad m = 1, 2, \ldots, L,$$

where $H \in R^{N\times N}$ is an lower triangular matrix with the term of $h_{mn} = h_{m-n}$, and $\{h(k)\}$ is FIR of $A^{-1}(z) = 1/(1 + \sum_{k=1}^{Q} a_k z^{-k})$. Here, $\mathbf{e}_m$ which is assumed to be $\mathbf{e}_m^T \mathbf{e}_m = constant$ (e.g., $constant = w^2 > 0$) for all m denotes excitation vector of the codeword $\mathbf{y}_m$ and α_m is the gain of $H\mathbf{e}_m$. The conditional MSE, which is denoted by D, between the desired vector $\mathbf{y} \in R^{N\times 1}$ (with its nth component given by $y(n)$) and the vector $\mathbf{y}_m$ is given by

$$D = E[\min_m ||H(\mathbf{e} - \alpha_m \mathbf{e}_m)||^2 \mid \mathbf{e}]$$

$$= \quad E[\min_m ||\mathbf{e} - \alpha_m \mathbf{e}_m||_W^2 \mid \mathbf{e}], \tag{2}$$

where $||\mathbf{x}||^2 = \mathbf{x}^T\mathbf{x}$, $||\mathbf{x}||_W^2 = \mathbf{x}^T W \mathbf{x}$, and $W = H^T H$.

MEAN SQUARED ERROR ANALYSIS

In general, if the size of stochastic codebook L becomes large, MSE becomes small. The length of frame N becomes large, it is expected that MSE is increased. Thus we observe that there is important relationship between MSE, N, and L. In this section, we investigate this relationship.

Observation 1: Suppose that

$$D = E[\min_m ||\mathbf{e} - \alpha_m \mathbf{e}_m||^2 \mid \mathbf{e}]. \tag{3}$$

Then we have an approximation of D which is given by

$$D \simeq ||\mathbf{e}||^2 Q(N, L) \tag{4}$$

where

$$Q(N, L) = \frac{(N-1)\Delta^2}{12}\left(1 - \frac{(N-1)\Delta^2}{4 \cdot 12}\right) \tag{5}$$

and

$$\Delta = \left(\frac{2\pi^{N/2}}{L\Gamma(N/2)}\right)^{\frac{1}{N-1}}. \tag{6}$$

Proof: See Appendix.

Let us consider more general case that

$$\begin{aligned} D &= E[\min_m ||\mathbf{y} - \mathbf{y}_m||^2 \mid \mathbf{e}] \\ &= E[\min_m ||\mathbf{e} - \alpha_m \mathbf{e}_m||_W^2 \mid \mathbf{e}], \end{aligned} \tag{7}$$

where $W \neq I$. In general, it is not easy to obtain D in (7), because we cannot use uniform pdf approximation to obtain distance between two vectors, $\mathbf{e}$ and $\mathbf{e}_k$, because $W \neq I$. We can, however, obtain D for special W, and the result is the following.

Observation 2: Suppose that $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ are the eigenvalues of the matrix W. Assume that $\lambda_1 \simeq \lambda_2 \simeq \cdots \simeq \lambda_M \gg \lambda_{M+1} \simeq \lambda_{M+2} \simeq \ldots \simeq \lambda_N$. Then we have an approximation that

$$\begin{aligned} D &= E[\min_m ||\mathbf{e} - \alpha_m \mathbf{e}_m||_W^2 \mid \mathbf{e}] \\ &\simeq ||\mathbf{e}||_W^2 Q(M, L). \end{aligned} \tag{8}$$

Proof: Suppose that $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N\}$ are the eigenvectors of W. Then we have

$$W = G\Lambda G^T,$$

where $\Lambda = diag[\lambda_1, \lambda_2, \ldots, \lambda_N]$ and $G = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N]$. Since G is an orthonormal matrix, geometrical properties between $G^T\mathbf{x}$ and $\mathbf{x}$ do not differ from each other. Thus we simply use Λ instead of W without loss of generality, i.e., we have

$$E[\min_m ||\mathbf{e} - \alpha_m \mathbf{e}_m||_W^2 \mid \mathbf{e}] = E[\min_m ||\mathbf{e} - \alpha_m \mathbf{e}_m||_\Lambda^2 \mid \mathbf{e}].$$

Let us denote that $\mathbf{e} = [e_1, e_2, \ldots, e_N]^T$ and $\mathbf{e}_m = [e_{m1}, e_{m2}, \ldots, e_{mN}]^T$, $m = 1, 2, \ldots, L$. Then we have

$$||\mathbf{e} - \alpha_m \mathbf{e}_m||_\Lambda^2 \simeq \lambda_1(e_1 - \alpha_m e_{m1})^2 + \lambda_2(e_2 - \alpha_m e_{m2})^2 + \cdots + \lambda_M(e_M - \alpha_m e_{mM})^2,$$

because $\lambda_1 \simeq \lambda_2 \simeq \cdots \simeq \lambda_M \gg \lambda_{M+1} \simeq \lambda_{M+2} \simeq \ldots \simeq \lambda_N$. In this case, only M elements of N elements of $\mathbf{e}$ and $\mathbf{e}_m$ play dominant role in determination of codeword. Thus we can use the result of **Observation 1** by changing N to M.

It is observed that dimension reduction has been occurred, i.e., M dimensions are emphasized by M largest eigenvalues and the other dimensions are not contributed, because corresponding eigenvalues are relatively quite smaller than the Mth largest eigenvalue, λ_M. Hence we can define the quantity *number of effective dimensions* M. Since the number of effective dimensions M is determined by eigenvalues of the matrix W, it depends on correlation properties of signal $y(n)$. Let us consider an extreme case that signal $y(n)$ is a constant for all n or quite correlated, then the matrix W becomes a rank one matrix. It implies the number effective dimensions $M = 1$. On the other hand, it is clear that if signal $y(n)$ are close to white process, the number of effective dimensions M becomes N. It is noteworthy that the number of effective dimensions M is closely related to the power spectrum of $y(n)$.

Let us now consider the case that $\lambda_M \gg \lambda_{M+1}$ is not satisfied, i.e., we cannot find M significantly large eigenvalues. In this case, it is not easy to determine the number of M. In the next section, we consider a method to determine the number of M when the condition $\lambda_M \gg \lambda_{M+1}$ is not satisfied.

DETERMINATION OF NUMBER OF EFFECTIVE DIMENSIONS M

Let us consider determination of the number of M of the matrix W. When we cannot find significantly large M eigenvalues of W, it is not easy to determine the number of effective dimensions M of N. In section, we consider a method which is based on threshold approach to determine the number of effective dimensions M.

One of easy method to determine the number of effective dimensions M is threshold approach. Let us define threshold T and consider a method which can determine the number of effective dimensions M:

$$M = \arg\min_P \{\sum_{k=1}^{P} \lambda_k \geq T \sum_{k=1}^{N} \lambda_k\}, \tag{9}$$

where $0 < T < 1$ and the value of T should be close to 1. With this approach, threshold T should be determined in a priori.

APPLICATIONS OF MSE FORMULA

In this section, we consider two applications of Eq. (8). One application is for adaptive size of stochastic codebook and the other is for stopping rule to reduce codebook search time.

Adaptive size of stochastic codebook

For large number of effective dimensions, the improvement by increasing the size of stochastic codebook is not significant. For a frame of which the number of effective dimensions is large, the size of stochastic codebook can be decreased without significant performance degradation. From this observation, both codebook search time and number of bits for encoding of codewords can be reduced using adaptive size of stochastic codebook depending on the number of effective dimensions for each frame.

Stopping rule for stochastic codebook search

Let us now consider stochastic codebook search of CELP coder. When the size of stochastic codebook is L, to find the best codeword, L iterations should be performed. Utilizing (8), we can make a stopping rule which can be used for reducing codebook search time.

We now briefly consider a stopping rule which is based on (8): while stochastic codebook is searched, if the squared error of the kth codeword is smaller than given threshold $T(D)$ which is a function of D in (8), we use the kth codeword to excitation vector and stop iteration. Of course, this stopping rule does not provide the best codeword, however, squared error can be close to MSE D when $T(D) = D$. If one wants to reduce codebook search time with relatively small increment of squared error, larger than MSE D, threshold $T(D)$ can be slightly greater than D. Some examples of implementation of the stopping rule in CELP coder are shown in the next section.

SIMULATION RESULTS

In this simulation, we use CELP coder with frame of length $N = 60$, linear predictor of order $Q = 10$, and the self-excitation codebook of length 128. The stochastic codebook contains sparse, overlapping, and ternaly valued pseudo randomly generated codewords. This CELP coder is quite similar to, but simpler than DoD CELP coder [1]. With this CELP coder the two applications in Section 5 have been applied to reduce stochastic codebook search time.

In Figure 1, with this CELP coder we obtain a sequence of squared errors for each frame. The solid and dotted

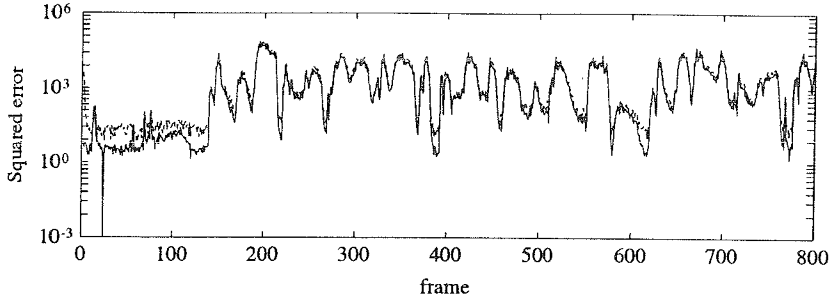

Figure 1: Sequences of theoretical approximation of and actual squared errors of a speech signal.

lines represent theoretical approximations of and actual squared errors, respectively. It is seen that theoretical approximation of squared error is smaller than actual squared error. This comes from that the stochastic codebook of the CELP coder contains overlapping and ternaly valued codewords of stochastic codebook. Thus actual squared error can be larger than that of theoretical approximation.

When we apply adaptive size of stochastic codebook, the stochastic codebook search time is reduced to approximately 41% of full search time of stochastic codebook with sacrifice of increment of squared error by 1.063 time. Moreover, we can have over 25% reduction of number of bits for encoding codewords of stochastic codebook.

CONCLUDING REMARKS

Mean squared error of linear prediction coder with excitation from stochastic codebook has been considered in this paper. It is seen that mean squared error depends on both size of stochastic codebook and length of frame. Moreover, it is observed that the correlation properties of desired signal affect mean squared error. It is often observed that when desired signal is close to uncorrelated signal, mean squared error becomes large. On the other hand, desired signal is quite correlated signal, mean squared error becomes small. Using the number of effective dimensions, the relationship between mean squared error and correlation properties of desired signal can be obtained. When desired signal is close to uncorrelated signal, it is shown that number of effective dimensions becomes large and it yields large value of mean squared error. When desired signal is quite correlated signal, it has small number of effective dimensions and it leads small value of mean squared error.

In this paper, we use threshold approach to determine number of effective dimensions. Other parametric and model based approach, however, are desired and should be considered as further study. We only consider analysis of mean squared error in terms of stochastic codebook. Analysis of mean squared error in terms of the codebook of self-excitation as well as of vector sum excited linear prediction coder are highly desired and considered as further study.

Two applications of mean squared error formula have been considered to reduce stochastic codebook search time with sacrifice of small increment of squared error in CELP coder. One of two can be used for reduction of number of bits to encode codewords of stochastic codebook.

APPENDIX

Part A

Let us assume that

$$\mathbf{e}_m^T \mathbf{e}_m = w^2, \tag{10}$$

and define

$$s^2 = ||\mathbf{e}||^2.$$

It is noteworthy that the vectors $\{\mathbf{e}_m\}$ can be represented by points on the hypersphere of radius w in N-dimensional vector space. Using s^2 and w^2, the vectors $\mathbf{e}$ and $\mathbf{e}_m$ can be normalized which are given by

$$\bar{\mathbf{e}} = \frac{1}{s}\mathbf{e} \quad \text{and} \quad \bar{\mathbf{e}}_m = \frac{1}{w}\mathbf{e}_m,$$

then we have

$$\min_m \|\mathbf{e} - \alpha_m \mathbf{e}_m\|^2 = s^2 \min_m (1 - (\bar{\mathbf{e}}^T \bar{\mathbf{e}}_m)^2),$$

where the gain is determined by

$$\alpha_m = \frac{\mathbf{e}^T \mathbf{e}_m}{\mathbf{e}_m^T \mathbf{e}_m}$$

to minimize $\|\mathbf{e} - \alpha_m \mathbf{e}_m\|^2$ itself.

Suppose that $\bar{\mathbf{e}} \simeq \bar{\mathbf{e}}_k$ and

$$\bar{\mathbf{e}} = \bar{\mathbf{e}}_k + \mathbf{v}_k. \tag{11}$$

Since $\|\bar{\mathbf{e}}\|^2 = \|\bar{\mathbf{e}}_k\|^2 = 1$, we have

$$2\mathbf{v}_k^T \bar{\mathbf{e}}_k = -\mathbf{v}_k^T \mathbf{v}_k \tag{12}$$

and

$$E[\min_m \{1 - (\bar{\mathbf{e}}^T \bar{\mathbf{e}}_m)^2\} \mid \bar{\mathbf{e}} = \bar{\mathbf{e}}_k + \mathbf{v}_k] = E[\|\mathbf{v}_k\|^2] - \frac{1}{4}E[\|\mathbf{v}_k\|^4]. \tag{13}$$

It is noteworthy that the random vector $\mathbf{v}_k$ can be represented by a vector in $(N-1)$-dimensional vector space from (12) and can be approximated by a vector in the hypersphere of radius 1 in N-dimensional vector space, i.e., we can have

$$\mathbf{v}_k^T \mathbf{v}_k = \sum_{l=1}^{N-1} \xi_l^2, \tag{14}$$

where $\{\xi_l\}$ are independent random variables.

If L is sufficiently large and $\{\bar{\mathbf{e}}_m\}$ are placed with equi-distance over the hypersphere of radius 1 in N-dimensional vector space, we can assume that $\{\xi_l\}$ are random variables which are independent each other and of common uniform pdf $U[-\Delta/2, \Delta/2]$, where determination of the value of Δ is in Part B. Thus we have

$$\begin{aligned} E[\mathbf{v}_k^T \mathbf{v}_k] &= E[\sum_{l=1}^{N-1} \xi_l^2] \\ &= (N-1)\frac{\Delta^2}{12} \end{aligned} \tag{15}$$

and

$$\begin{aligned} E[(\mathbf{v}_k^T \mathbf{v}_k)^2] &= E[(\sum_{l=1}^{N-1} \xi_l^2)^2] \\ &\simeq [(N-1)\frac{\Delta^2}{12}]^2. \end{aligned} \tag{16}$$

The approximation in (16) comes from that $E[\xi^4] = \Delta^4/80 \simeq (\Delta^2/12)^2$.

Consequently, using (13), (15), and (16), we have

$$D = \|\mathbf{e}\|^2 \frac{(N-1)\Delta^2}{12}\left(1 - \frac{(N-1)\Delta^2}{4 \cdot 12}\right). \tag{17}$$

Part B: Determination of Δ

Let us denote S the area of the hypersphere of radius 1 in N-dimensional vector space, it is given by [4]

$$S = \frac{2\pi^{N/2}}{\Gamma(N/2)}.$$

Since we assume that the codewords are placed with equi-distance on the hypersphere, we can determine Δ which is given by

$$L\Delta^{N-1} = S$$

or

$$
\begin{aligned}
\Delta &= \left(\frac{S}{L}\right)^{\frac{1}{N-1}} \\
&= \left(\frac{2\pi^{N/2}}{L\Gamma(N/2)}\right)^{\frac{1}{N-1}}
\end{aligned}
$$

References

[1] J.P. Campbell, T.E. Tremain, and V.C. Welch, "The DoD 4.8 Kbps standard (proposed federal standard 1016)," in *Advances in Speech Coding*, pp. 121-133, B.S. Atal, V. Cuperman, and A. Gersho (eds), Kluwer Academic Pub., Norwell, Massachusetts, 1991.

[2] J.-H. Chen, R.V. Cox, Y.-C. Lin, N. Jayant, and M.J. Melchner, "A low-delay CELP coder for the CCITT 16kb/s speech coding standard," *IEEE Jour. Sel. Areas in Comm.*, vol. SAC-10, pp. 830-849, June 1992.

[3] C.R. Galand, J.E. Menez, and M.M. Rosso, "Adaptive code excited predictive coding," *IEEE Trans. Signal Proc.*, vol. SP-40, pp. 1317-1326, June 1992.

[4] A. Gersho and R.M Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Pub., Norwell, Massachusetts, 1992.

[5] I. Gerson and M.A. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8kbps," in *Proc. IEEE Int. Conf. Acous., Speech, Signal Proc.*, pp. 461-464, April 1990.

[6] I.M. Trancoso and B.S. Atal, "Effective search procedures for selecting the optimum innovation in stochastic coders," *IEEE Trans. Acous., Speech, Signal Proc.*, vol. ASSP-38, pp. 385-396, March 1990.