2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

A MIXED GAUSSIAN-STOCHASTIC CODE BOOK FOR CELP CODER IN LSP SPEECH CODING

Najib Naja, Jean Marc Boucher and Samir Saoudi

Groupe Communications Numériques
Département Mathématiques et Systèmes de Communications
ENST-Br. BP 832 - 29285 Brest - FRANCE

10.21437/ICSLP.1992-409

ABSTRACT

In a last few years, the Code Excited Linear Predictive (CELP) coder has demonstrated that such coders offer considerable promise for producing high quality synthetic speech at bit rates between 4.8 and 16 kbit/s. Tremendous efforts have been made to improve the CELP coder performance or to reduce its implementation cost. The CELP performance has been studied for different kind of code books such as Gaussian code books, algebraic code books, ternary code books ...

In this paper, the Gaussian and a mixed Gaussian-stochastic code books were compared at different sizes, while the Line Spectrum Pairs (LSP) were used instead of Log Area Ratios (LAR) which provide efficient representation of the synthesis filter used in the CELP. This new stochasic code book was designed for various sizes, using the K-means algorithm applied to a large training data. Its first half is a Gaussian code book and the other half is obtained from the prediction error (after standardizing to unit variance) of speech signals when the skewness β_1 and the kurtosis β_2 parameters do not correspond to the Gaussian law. For the same size, the new stochastic code book offers a better performance than the Gaussian one, about 0.5 dB for the same number of bits.

I. INTRODUCTION

Several analysis by synthesis coders have been developed in the rate range from 4.8 to 16 kbit/s [7]. These systems include the Code Excited LPC (CELP) [13], the self-excited LPC [11] and the regular-pulse excited LPC (RPE-LPC) [6].

The CELP has proven to be the most promising candidate for producing high quality speech at bit rates ranging from 8 to 4.8 kbit/s, where bit rate reduction is achived by vector quantizing the excitation sequence using a large stochastic code book. Atal [2] has show that the probability density function (pdf) of the prediction error samples (after both short-delay and long-delay predictions) is nearly Gaussian. A large Gaussian code book (usually 1024 entries) is used and the optimum innovation sequence is determined by exhaustively searching the code book for the address (and the corresponding gain) which minimizes the mean-squared weighted error criterion. However, the CELP complexity has been reduced by using efficient code books structures such as algebraic code books [1], [8], ternary code books [9], overlapping sparse code books [5] [9] and vector sum excitation [4].

Schroeder and Atal [13] have proved that the Gaussian assumption is valid almost everywhere except for stop bursts of unvoiced stop consonants and for a few pitch periods during the transition from unvoiced or silence regions to voiced speech. This hypothesis is clearly true for 80 % of short speech blocks of 2.5 msec duration, which is shown by computing skewness β_1 and kurtosis β_2. For the remaining 20 % there is a deviation from the previous statistic.

The aim of this work is to design a new stochastic code book which fits better with the statistics. It uses K-means algorithm [10] and has been applied to great number of innovation sequences. Its first half is a Gaussian code book and the other half is obtained from the prediction error (after standardizing to unit variance) of speech signals when the β_2 and / or β_1 parameters do not correspond to the Gaussian law.

II. DESCRIPTION OF THE CELP STRUCTURE

The basic CELP coder as shown by Fig. 1 use only analysis by synthesis procedures for determinig the optimal innovation sequence from a stochastic code book.

With the decoder structure of Fig. 1-b, the synthesized speech s(n) is obtained by filtering the excitation U(n) with a long term synthesis filter of gain b and delay P and the all pole LPC filter 1/A(z) in cascade.

The coder structure of Fig. 1-a was obtained in [8] with an analysis by synthesis procedure [14] for selecting the optimum excitation from a set of N waveforms of length L $\{U_i(n), n=0,\ldots,L-1; i=0,\ldots,N-1\}$ scaled by a gain factor G(i). The main entity of the scheme is e(n), the input residual signal to be produced. The linear prediction residual e(n) results from filtering the original signal s(n) by A(z). The gain b and the delay P of the long term predictor are then evaluated via the autocorrelation function of e(n).

The residual signal d(n) or difference signal is obtained by substracting from e(n) the signal $\hat{e}(n)$ equal to the reconstructed residual signal $\tilde{e}(n)$ for n < 0, and to the long term prediction of the reconstructed residual signal for n=0,...,L-1:

$$d(n) = e(n) - \hat{e}(n),$$

$$\text{with} \quad \hat{e}(n) = \begin{cases} \tilde{e}(n) & n < 0 \\ b\,\tilde{e}(n-P) & n = 0,\ldots,L-1 \end{cases} \qquad (1)$$

Now let us define the weighted input signal $e_w(n)$ and the weighted excitation $U_{wi}(n)$ by:

$$e_w(n) = d(n) * hg(n), \qquad (2)$$

$$U_{wi}(n) = U_i(n) * hg(n) \quad i = 0,\ldots,L-1, \qquad (3)$$

where hg(n) is the impulse response of the weighting filter 1/Ag(z). Ag(z) is obtained from A(z) by replacing z by $g^{-1}z$. A FIR filter can be used to evaluate $e_w(n)$. As typical values of g are close to 0.8 the impulse response hg(n) of the perceptuel filter 1/Ag(z) will be of moderate lenght say K = 15.

Finally the best waveform I and the gain G(I) are found by minimizing the following error criterion:

$$D(i) = \sum_{n=0}^{L-1} \left[e_w(n) - G(i) U_{wi}(n) \right]^2 \quad i = 0, ..., N-1 \quad (4)$$

The index I and the gain G(I) are transmitted to the decoder. The reconstructed residuel signal $\tilde{e}(n)$ used for the next block in (1) is obtained by adding the optimum excitation to the long term prediction [3].

III. CODE BOOK GENERATION

To design the code book, we partition the L-dimensional space of the input vectors x of the training sequence into M clusters or cells $\{S_i, i=0,...,M-1\}$ and associate each cell S_i with a vector c_i, called the codeword or centroid. Each codeword is characterized by its index i.

Let R be the number of bits necessary to code the index, so we have

$$M = 2^R \quad (5)$$

In this paper R is varying from 5 to 11.

Taking as reference the K-means[10] clustering algorithm, we develop the following algorithm for code book generation, where we prevent empty cells:

Step 0:
Initialization: Close an initial code book. Here,the initial code book is populated with the M first summits of the L-dimensional cube,$\{x:|x_i|<4;i=0,...,L-1\}$ including all or most of the points in the training sequence.

Step 1:
Classification: Classifay the training sequence into the clusters S_i using a minimum distortion rule. Compute the average distortion D.

Step 2:
Updating: Update the centroid of every cluster. If any cluster S_j is empty, we assign to it the j^{th} centroid from the previous iteration.

Step 3:
Convergence test: Compute $(D_{old} - D)/D$ and test if it is below a certain threshold ε (ε=0.001). If so stop. Otherwise go to step 1.

As a distortion measure, we use the squared-error distortion:

$$d(x,c) = \sum_{j=1}^{L} (x_j - c_j)^2 \quad (6)$$

where x and c denote a points from the training sequence and the codeword respectively.

IV. MEASURES OF SKEWNESS AND KURTOSIS

For a symetrical distribution, mean, median and mode coincide. It is thus natural to take the deviation mean to mode or mean to median as measuring the skewness of the distribution. K. Pearson proposed the measure

$$Sk = \frac{\text{mean} - \text{mode}}{\sigma}$$

which is subject to the inconvenience of determinig the mode. For a wide class of frequency-distributions known as Pearson's, this measure may, however, be expressed eaxctly in terms of the first four moments. We define [15]

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad (7)$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad (8)$$

$$\text{with} \quad \mu_\alpha = E[(X - E(X))^\alpha] \quad (9)$$

The coefficient $\beta 1$ is a measure of skewness. The $\beta 2$ parameter is a measure of kurtosis.

In the normal distribution (Gaussian law) we have:

$$\beta 1 = 0 \quad \text{and} \quad \beta 2 = 3.$$

We can then conclude if the probability law to which a random signal obey is not Gaussian by computing the deviations of its skewness $\beta 1$ and its kurtosis $\beta 2$ from the previous values.

V. EXPERIMENTAL RESULTS

The CELP system described in Fig.1 uses a 10^{th} order short term predictor. The LSP [12] parameters are updated every 20 msec, the long-term predictor parameters every 10 msec, and the code word and gain term every 2.5 msec. A Hamming window of 28 msec is used to weight a speech data.

5.1 Training sequence design

The training corpus for the design of the training sequence consists of 210 double sentences sampled at 8 kH (10 speakers : 5 male and 5 female) i.e. 22mn.

As a first experiment, we determine if the prediction error samples e(n) (after both short-delay and long-delay prediction) has a Gaussian probability law or not. Aln this aim, we compute $\beta 1$ and $\beta 2$. We also compute the means $\beta 1_{,G}$ and $\beta 2_{,G}$ and the standard deviations $\sigma_{\beta 1,G}$ and $\sigma_{\beta 2,G}$ for a large Gaussian set of blocks of 2.5 msec. duration, to serve as reference. This can be justified by the fact that we compute $\beta 1$ and $\beta 2$ for a small number of samples which introduce an error for their estimations. Comparing the $\beta 1$ and $\beta 2$ parameters estimated for each block of 20 consecutive error e(n) to the empirical values $\beta 1_{,G}$ and $\beta 2_{,G}$ ($\beta 1_{,G} = 0.22$ and $\beta 2_{,G} = 2.72$) can reduce the number of our error decision concerning the Gaussianity of the probability law of the error e(n). The probability density function of the prediction error samples is nearly Gaussian when the computed $\beta 1$ and $\beta 2$ parameters differ less than $\sigma_{\beta 1,G}$ and $\sigma_{\beta 2,G}$ ($\sigma_{\beta 1,G} = 0.33$ and $\sigma_{\beta 2,G} = 0.73$) from $\beta 1_{,G}$ and $\beta 2_{,G}$ respectively. This hypothesis is clearly true for 80% of short speech blocks 2.5 msec duration. The 20% remaining (Fig. 2) constitue our training sequence (about 138,000 innovation sequences after standardizing the prediction error to unit variance).

5.2 The code book design

For the second experiment, our mixed Gaussian-stochastic code book was designed for various sizes, using the K-means algorithm (section III) applied to the training sequence established as explained in 5.1 to determine a half of it. The other half is a Gaussian code book.

5.3 Objective performance

For the same size, the Gaussian code book and this new stochastic code book are compared, when the LSP parameters are quantized by a number B of bits per frame varying from 20 to 40. The LSP parameters are quantized by an optimal quantization method [13].

The curves of global and segmental signal to ratios versus the number of bits with the quantization of the LSP are sketched for the two kinds of code books at different sizes in Fig.3 (a,b,c). The new stochastic code book offers a better performance than the Gaussian one. For B > 28, we can use a mixed Gaussian-stochastic code book in order to use a Gaussian one with twice the size to have roughly the same speech quality.

VI. CONCLUSIONS

We have shown that in the case of a CELP with LSP parameters, the innovation sequences corresponding to a non-Gaussian probability law can serve to design a mixed Gaussian-stochastic code book based on the K-means. This new code book and a Gaussian one are compared at the same sizes via the global and segmental signal to residual ratio. It is shown that the new code book has better performances than the Gaussian one.

REFERENCES

[1] J-P Adoul et al," Fast CELP coding based on albebraic codes ," Proc.ICASSP'87, pp. 1957-1960.
[2] B. S. Atal," Predictive coding of speech at low bit rates," IEEE Trans. Commun. vol. COM-30, pp. 600-614, April 1982.
[3] C. Galand, M. Rosso, P. Elie, E. Lançon," MPE/LTP speech coder for mobile radio application". Speech Communication, Vol. 7-2, 1988, pp. 167-178.
[4] I. A. Gerson and M. A. Jaiuk,"Vector Sum Excited Linear Prediction (VSELP)," IEEE workshop on speech coding for Telecom., Sept. 5-8 1989, Vancouver, Canada.
[5] W. B. Kleijn, D. J. Krasinsky and R. H. Ketchum, "An efficient stochastically excited linear predictive coding algorithm for high quality low bit rate transmission of speech," Speech Commun., Vol. 7, no. 3, pp. 305-316, Oct. 1988.
[6] P. Kroon, E. F. Deprettere, and R. J. Sluyter, "Regular-pulse excitation - A novel approach to effective efficient multipulse coding of speech," IEEE Trans. ASSP, Vol. 34, no. 5, pp. 1054-1063, Oct. 1986.
[7] P. Kroon, E. F. Deprettere, "A class of analysis by synthesis predictive coders for high quality coding at rates between 4.8 and 16 kbit/s," IEEE J. Selected Areas in Commun., Vol. 6, no. 2, pp. 353-363, Feb. 1988.
[8] A. Le Guyader, "An efficient CELP structure for speech coding between 8 kbit/s and 16 kbit/s," Proc. EUSIPCO signal processing IV: Theories and applications, Elsevier Publishers B. V (North Holland) 1988, pp. 867-870.

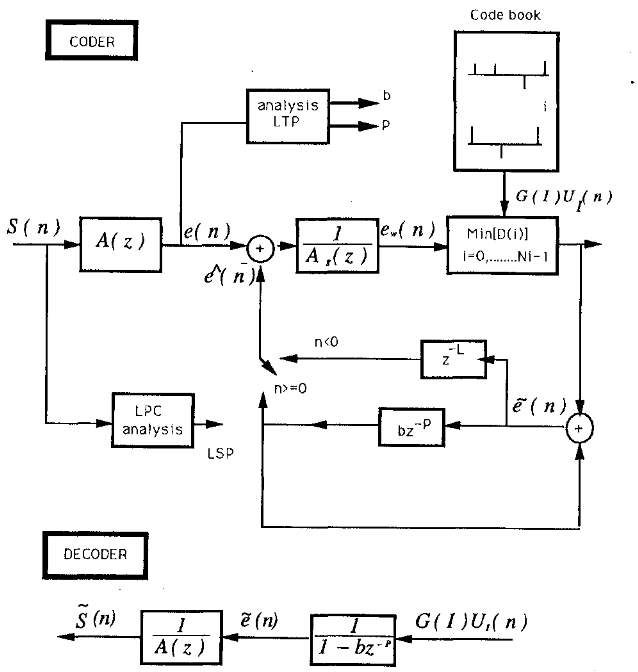

Fig.1 Principle of the CELP coder

[9] D. Lin, "New approaches to stochastic coding of speech sources at very low bit rates," Signal Processing III: Theories and Applications (Proc. of EUSIPCO-86), pp. 445-448, 1986.
[10] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., pp. 84-95, 1980.
[11] R. C. Rose and T. B. barnwell III, "Quality comparison of low complexity 4800 bps self excited and code excited vocoders," Proc. ICASSP 1986, pp. 2375-2378.
[12] S. Saito and K. Nakata, "Line Spectrum Pairs," Fundamentals of speech signal processing, Academic press, Japan, 1985, chap. 9, pp.127-132.
[13] S. Saoudi, J.M. Boucher and A. Le Guyader, "Optimal scalar quantization of the LSP and the LAR for speech coding", ICSLP-90; Kobe, pp.4.4.1-4.4.4., November 1990.
[14] M. R. Schroder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High-quality speech at very low bit rates," Proc. ICASSP'85, pp. 937-940.
[15] A. Stuart and J. K. Ord, "Kendall's Advanced Theory of Statistics," Volume 1: Distribution Theory, Charles Griffin & Company Limited, pp. 106-110.

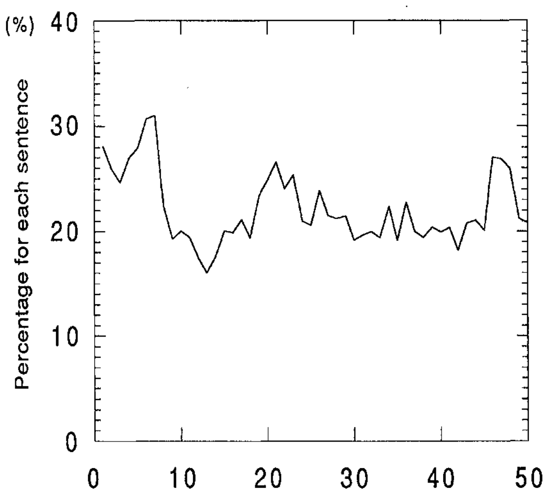

Fig. 2 Curve of non-Gaussian blocks

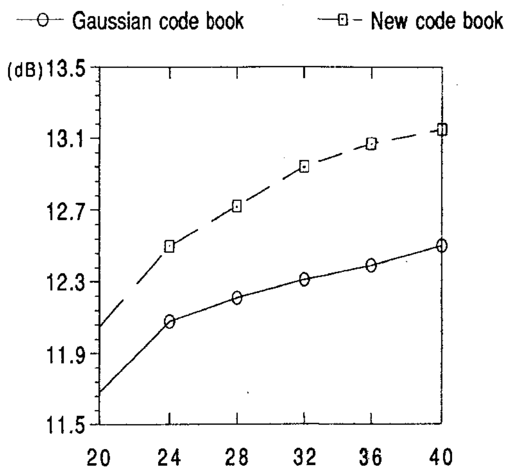

Fig. 3 (a) RSB versus the bit rate per frame with LSP (code book size = 4096)

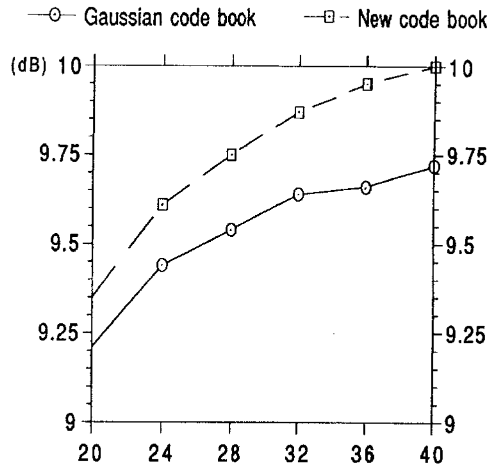

Fig. 3 (a) RSBseg versus the bit rate per frame with LSP (code book size = 4096)

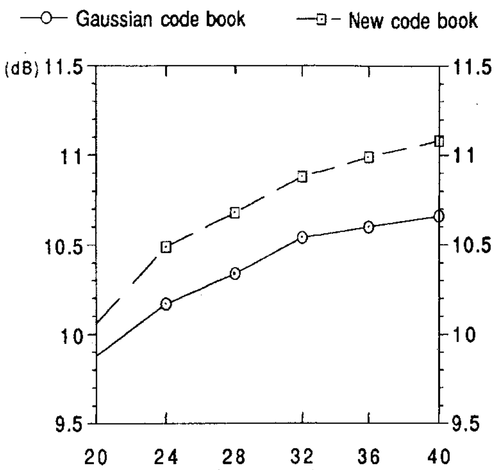

Fig. 3 (b) RSB versus the bit rate per frame with LSP (code book size = 512)

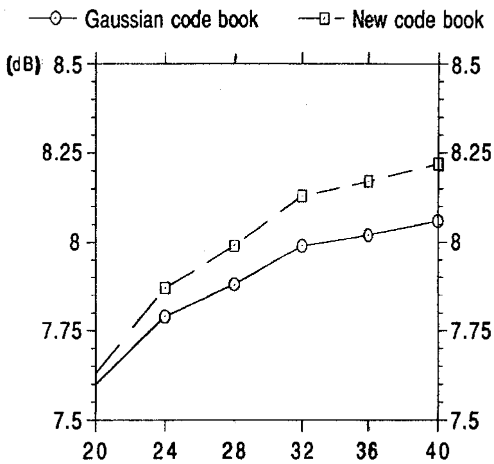

Fig. 3 (b) RSBseg versus the bit rate per frame with LSP (code book size = 512)

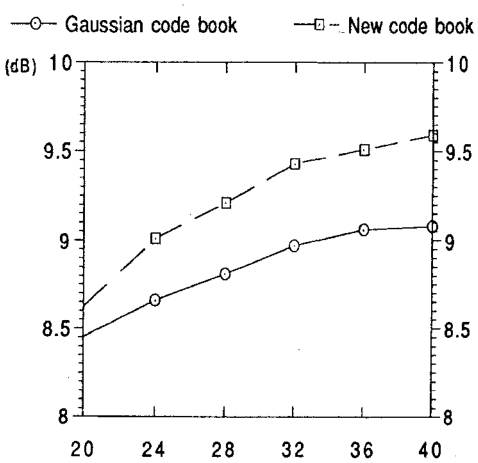

Fig. 3 (c) RSB versus the bit rate per frame with LSP (code book size = 128)

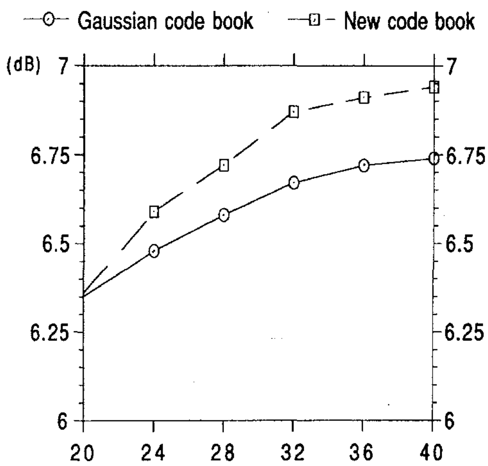

Fig. 3 (c) RSBseg versus the bit rate per frame with LSP (code book size = 128)