Dynamic partial search scheme for stochastic codebook of FS1016 CELP coder

A.-N. Suen
J.-F. Wang
T.-C. Yao

Indexing term: Speech coding algorithms

Abstract: The authors present a new classified dynamic partial search structure for the stochastic codebook of the FS1016 CELP coder to replace the fixed partial search for selecting the best excitation vector of the stochastic codebook. In the proposed scheme, the conventional one-stage stochastic codebook search is substituted with a two-stage dynamic method for reducing the computational complexity without degrading the voice quality. The establishment of this structure is based on two classifiers, one for the line spectrum pairs (LSP) of the input signals, and the other for the autocorrelation coefficients (AC) of the stochastic codebook search target. In addition, the stochastic codebook is classified into K subcodebooks, and with these two classifiers it is possible to determine dynamically which subcodebook needs to be searched. This method achieves a reduction in the search procedure by a factor of 2–8. The efficiency of these two classifiers is discussed and the comparison of the performance between the fixed partial search and the proposed technique is also addressed.

1 Introduction

Many potential applications for low bit rate speech coding algorithms demand good speech quality at a reasonable cost. The ability to encode the speech at low bit rates without destroying voice quality is becoming increasingly important in the new digital communication environment.

The code excited linear predictive (CELP) coder jointly developed by the DoD and AT&T Bell Laboratories [1] has been proposed as the US Government standard 4.8 kbit/s voice coder [2]. The computational requirement of CELP is dominated by the adaptive and stochastic codebook searches. The full stochastic codebook search procedure has a heavy computational load due to the convolution operations between the excitation vectors and the linear predictor filter. The fixed partial search scheme is the most popular method to reduce the computational load, but the speech quality will be degraded when using such a partial search scheme. The recent development of fast methods for the CELP coder [3–5] have provided potential algorithms to reduce the computational complexity of codebook searches. But the cost is still high if the exhaustive search is adopted. Performance constrained algorithms such as tree structure [6], finite state structure [7, 8], neural network [9, 10] and the preselection schemes [11–13] are also proposed for reducing the coding time. The above performance constrained strategies are hard to apply directly to the FS1016 CELP coder because they are all constructed based on the special codebook structures. In Reference 14, Mauc mentions a multistage method for eliminating the irrelevant codevectors of the stochastic codebook of FS1016.

In this paper we present a dynamic partial stochastic codebook search (DPSCS) method for the FS1016 CELP coder. The emphasis of this work is on the efficient search procedure of the stochastic codebook with two classifiers. Our approach would not only reduce the computational effort but also have a high hit rate for the optimal codeword without degrading the voice quality. The proposed dynamic partial search method is based on two classifiers, one for the line spectrum pairs (LSP) of the input signals and the other for the autocorrelation coefficients (AC) of the target signals of the stochastic codebook. The target signal of the stochastic codebook search is the weighted linear prediction residual plus the encoding error and minus the filtered adaptive codebook VQ excitation [2] (described in Section 2). In addition, the whole stochastic codebook is first classified into K subcodebooks with S codewords and then which subcodebook needs to be searched is determined by these two criteria parameters. In our DPSCS system, the adaptive codebook search is the same as the conventional FS1016 CELP coder, but the stochastic codebook search procedure has been replaced with a two-stage dynamic partial search scheme. The merits of the proposed scheme are (1) less extra memory required, (2) a higher hit rate for the optimal codevector, (3) reducing the complexity by a factor of 2–8 and (4) no degrading of the synthetic speech quality.

This paper is organised as follows. Section 2 briefly gives an overview of the conventional CELP coder algorithm. Section 3 describes the proposed DPSCS structure, including the establishment and incrementally updating procedures. Section 4 discusses the efficiency of the classifiers and shows the experimental results. Finally, we present our conclusions in Section 5.

This work was supported by the National Science Council. Republic of China, under Grant NSC84-2622-E-006-006.

Paper 1688K (E5), first received 6th May and in revised form 16th September 1994

The authors are with the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, ROC

2 CELP coder

Like all vector quantisation techniques, CELP coding is a frame-oriented technique that breaks the sampled input signal into blocks of samples that are processed as one unit. CELP coding is based on analysis-by-synthesis search procedures, perceptually weighted vector quantisation (VQ) and linear prediction (LP). The block diagram of the CELP coder is shown in Fig. 1. A 10th-order LP filter is used to model the short-term spectrum of the speech signal, or formant structure. Long-term signal periodicity, or pitch, is modelled by an adaptive codebook VQ. The residual from the short-term LP and pitch VQ is vector-quantised using a fixed stochastic codebook. The optimal scaled excitation vectors from the adaptive and stochastic codebooks are selected by minimising a time-varying, perceptually weighted distortion measure that improves subjective speech quality by exploiting masking properties of human hearing.

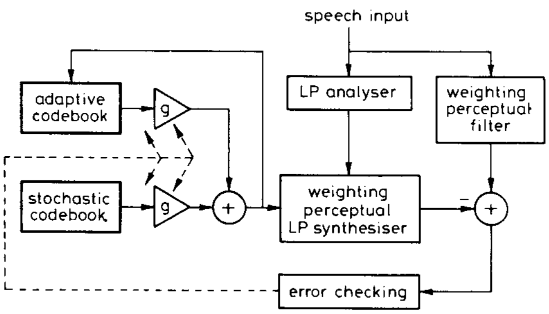

Fig. 1 *Block diagram of CELP coder*

This class of coders operates on sampled speech on a frame-by-frame basis. A filter is used to describe the spectral envelope of the speech signal. The coefficients of the filter are obtained using the linear prediction (LP) technique. They are quantised so that the same filter can be constructed at both the transmission and reception ends of the channel. The excitation for the filter is determined using an analysis-by-synthesis procedure [15]. A set of candidate excitation sequences is stored in a codebook, and synthetic speech is generated using each of these sequences. The index of the sequence producing the most accurate speech is then transmitted to the reception end of the channel.

Typically, the search for the optimal codebook vectors is done sequentially. The adaptive codebook is searched first, followed by a search of the fixed stochastic codebook. For each codebook, the optimal gain coefficient can be computed and, after all codebook indices have been found, the gain coefficients can be recomputed by solving a set of linear equations.

The error weighting filter $W(Z)$ is based on the short-term predictor $A(Z)$ with p coefficients:

$$A(Z) = 1 - \sum_{k=1}^{p} a(k)Z^{-k} \tag{1}$$

$$W(Z) = \frac{A(Z)}{A(Z/\gamma)} \tag{2}$$

where γ is typically 0.8.

Let $\boldsymbol{H}$ and $\boldsymbol{W}$ be $L \times L$ lower-triangular matrices whose columns contain the truncated impulse of the LP filter and error weighting filter, respectively, excited by a unit impulse on the diagonal:

$$\boldsymbol{H} = \begin{bmatrix} h_0 & 0 & 0 & \cdots & 0 \\ h_1 & h_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{L-1} & h_{L-2} & h_{L-3} & \cdots & h_0 \end{bmatrix}$$

$$\boldsymbol{W} = \begin{bmatrix} w_0 & 0 & 0 & \cdots & 0 \\ w_1 & w_0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{L-1} & w_{L-2} & w_{L-3} & \cdots & w_0 \end{bmatrix}$$

The adaptive codebook search target [2] can be represented as

$$T_a = \boldsymbol{W}(S - \hat{S}^{(0)}) \tag{3}$$

where S is the original speech signal and $\hat{S}^{(0)}$ is zero input response. The optimal excitation vector will be selected with the MSPE criterion. Let v be the best codeword for the adaptive codebook search stage; then the stochastic codebook search target [2] can be written as

$$T_s = T_a - \boldsymbol{WH}v \tag{4}$$

The CELP coder's computational requirements are dominated by the two codebook searches. The computational complexity and speech quality of the coder depend on the search sizes of the codebooks. Any subset of either codebook can be searched to fit processor constraints, at the expense of speech quality.

3 DPSCS architecture

Fig. 2 shows the flow chart of our dynamic partial search architecture. The criteria parameters are the LSP of the current frame and the autocorrelation coefficients (AC) of the stochastic codebook search target T_s (see eqn. 4). As the LSP parameters are extracted from each frame of input speech, the LSP cluster number i is determined with the Euclidean distance measurement. After the adaptive codebook search procedure, the AC cluster j is

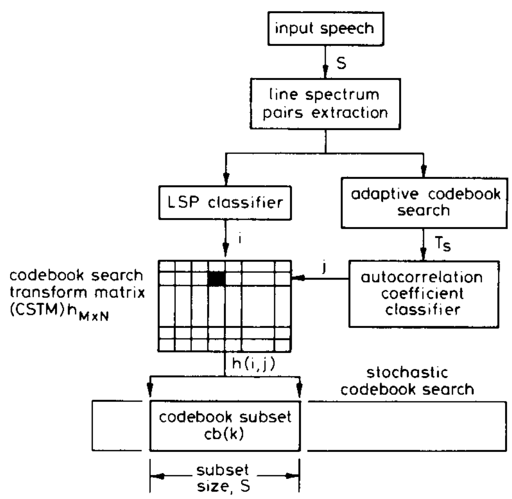

Fig. 2 *Flow chart of dynamic partial stochastic codebook search (DPSCS) scheme*

determined with the first-order AC of target signal T_s. With the LSP cluster i and AC cluster j, we can dynamically determine the partial codebook $cb(k)$ from the entry $h(i, j)$ of the codebook search transform matrix (CSTM). The CSTM contains the pointer which indicates the subset of the stochastic codebook that needs to be searched. The overall DPSCS system (shown in Fig. 3) includes the training and encoding phases. In the training phase, the LSP classifier and the CSTM will be constructed. The training and dynamic modification procedures of the DPSCS system will be described in detail as follows.

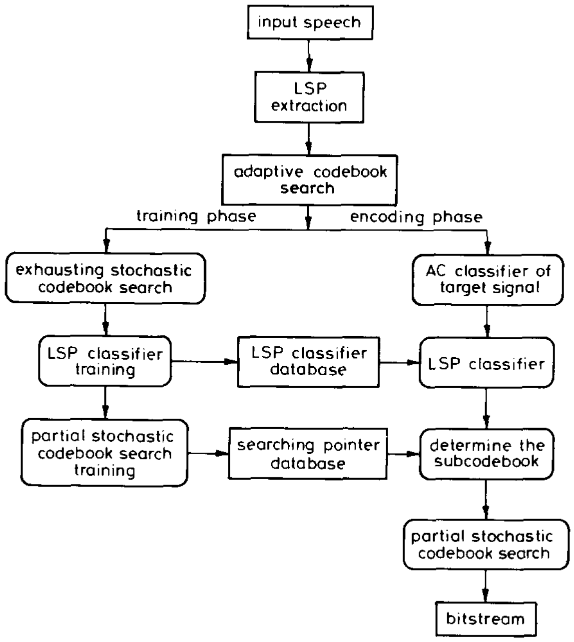

Fig. 3 *Overall system of DPSCS*

3.1 Autocorrelation classifier

The first stage of the proposed DPSCS scheme uses the AC of the stochastic codebook search target T_s to represent the waveform. There are two reasons for using the AC as the basis for classification. One is that AC can be used to classify the two different signals, and another is that the signal's AC will keep the same when scaled by a constant, and hence the gain's influence on the AC is disabled. To reduce the range of these coefficients distributed, the mth-order normalised AC, $a_x(m)$, is used and defined as eqn. 5:

$$a_x(m) = \frac{\sum_k (x_k - u)(x_{k+m} - u)}{\sum_k (x_k - u)^2} \tag{5}$$

where x is the input signal and u is the mean of the x values.

The experience distribution of $1 - a(1)$ is shown in Fig. 4, where $a(1)$ represents the 1st-order AC of the target signal. The γ distribution can be used to model the $1 - a(1)$ distribution, $(1 - a_x(1) \approx \gamma(x))$, and the γ distribution is defined as eqn. 6:

$$\gamma(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(\frac{-x}{\beta}\right) \tag{6}$$

It can be observed that the cumulative distribution of $1 - a(1)$ is very similar to the γ density function as shown in Fig. 5. The curve of the γ cumulative distribution can be divided into M equal areas, each area corresponding to one AC cluster. That is, each cluster contains the same $a(1)$ samples. Intuitively, multiorder AC classifiers may achieve better performance, but our experiments (described in Section 4) show that there is high correlation of $a(1)$ between the target and the optimal synthetic signals, but low correlation of $a(2)$ and $a(3)$. Therefore, only one AC classifier is used in our final DPSCS structure. While searching the optimal codeword in the stochastic codebook, we first calculate $a(1)$ of the target signal and then determine the cluster to which $a(1)$ belongs.

3.2 LSP classifier

The LSP analysis-synthesis method was first proposed by Itakura and Sugamura [16]. This method is treated as one of the most efficient speech analysis–synthesis techniques. Recently, there has been a growing interest in the use of the LSP parameters to code the short-time speech spectral [17, 18]. In the second stage of the proposed DPSCS scheme, we adopt the LSPs as the classifier criteria parameters. Initially, the LSP classifier with only one cluster is constructed. Determine the upper-bound LSP cluster number N, and the cluster number n ($n \leqslant N$) at the first training procedure. That is, we split the given LSP training sequence X into n partitions in the training phase (described in Step 3 of Section 3.3). Since it is impossible to collect all the LSP samples, the LSP classifier must be adaptive. If new training data come, the DPSCS just needs to employ the dynamic modification algorithm (described in Step 1 of Section 3.4) to modify the LSP classifier or increment one cluster.

The LBG algorithm [19] is employed to classify the LSP parameters with the Euclidean distance. The training process begins by calculating the Euclidean distance

d_{mq} between each cluster's centroid W_m and the input vector x_q according to

$$d_{mq} = \| W_m - x_q \|^2 = \sum_{p=1}^{P} (W_{mp} - x_{qp})^2 \quad (7)$$

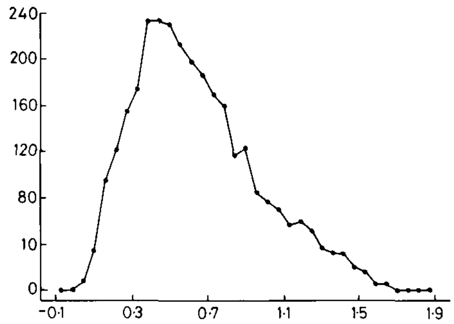

Fig. 4 *Histograms of $1 - a(1)$ of target signal*

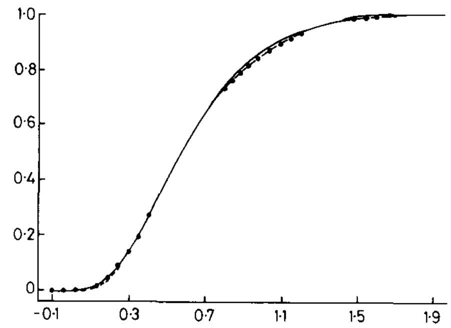

Fig. 5 *Cumulative distribution curves of $1 - a(1)$, and γ distribution with $\alpha = 3.64$ and $\beta = 5.722$*

where P is the dimension of the input vector.

This classifier architecture is similar to the memoryless one-stage vector-quantiser. By selecting the minimum Euclidean distance between the input LSP parameters x and each cluster's centroid, we can determine the LSP cluster to which the input LSP belongs.

3.3 Training of DPSCS

To obtain the dynamic partial stochastic codebook search (DPSCS) structure, two classified stages must be established first. For the LSP classifier stage, a simple vector quantiser with Euclidean distance is used. For the AC classifier stage, a scalar quantiser is used. The corresponding codewords for both quantisers are obtained from the training classified database.

With the above description, the algorithm for these two classifiers is shown as follows:

Step 1: Determine M, N and S, where M is the cluster number of AC, N is the upper bound of the LSP cluster number, and S is the subset size of the stochastic codebook.

Step 2: Determine the set of AC quantisation boundaries, $\{u(0), u(1), \ldots, u(M)\}$:

$$u(i) = \begin{cases} A + \left(i \dfrac{T}{M} - B\right) \dfrac{l}{t} & \text{for } i = 1, 2, \ldots, M - 1 \\ -1 & \text{for } i = 0 \\ 1 & \text{for } i = M \end{cases} \quad (8)$$

where A is the lower bound of cluster i, B is the total number of the classes whose ACs are smaller than A, and T represents the total AC samples. By sorting all the AC samples, we can find a range from A to $A + l$ containing t AC samples such that the numbers of ACs is as close to T/M as possible for each cluster. An alternative method of the AC classifying technique is the use of the γ distribution for modelling the AC distribution. The γ distribution curve can be divided into M parts with equal area, and each part indicates an AC cluster. The quantisation boundaries can be determined from a lookup table.

Step 3: Classify the given LSP training sequence X into n partitions $\{R(1), R(2), \ldots, R(n)\}$, and determine the corresponding centroid set $\{W(1), W(2), \ldots, W(n)\}$ and variance set $\{V(1), V(2), \ldots, V(n)\}$. The detailed training algorithm for classifying LSPs is described as follows:

(1) Define the training sequence $X = \{x_1, x_2, \ldots, x_m\}$, the cluster number n, $n \leqslant N$, and the distortion threshold $\varepsilon \geqslant 0$.

(2) Initially, only one cluster has been constructed: $\hat{n} = 1$, the partition $R(1) = X$, cluster centroid $W(1) = E(x \mid x \in R(1))$, and the variance $V(1) = E(\| x - W(1) \|^2 \mid x \in R(1))$, where E denotes the expectation.

(3) Split the cluster which has the maximum variance into two clusters: $\hat{n} = \hat{n} + 1$.

(4) Determine the partition set $\{R(1), R(2), \ldots, R(\hat{n})\}$:

$$R(j) = E(x \mid d(x, W(j)) \leqslant d(x, W(t)) \quad \text{for } j, t = 1, 2, \ldots, \hat{n} \quad (9)$$

where d indicates the Euclidean distance function.

(5) Record the cluster centroid $\hat{W} = W$ and determine the new cluster centroid set $\{W(1), W(2), \ldots, W(\hat{n})\}$:

$$W(j) = E(x \mid x \in R(j)) \quad \text{for } j = 1, 2, \ldots, \hat{n} \quad (10)$$

(6) Repeat (4) and (5) until the distortion $(D_{\hat{n}-1} - D_{\hat{n}})/D(\hat{n}) \leqslant \varepsilon$, where

$$D_{\hat{n}} = \frac{1}{\hat{n}} d(W(j) - \hat{W}(j)) \quad \text{for } j = 1, 2, \ldots, \hat{n} \quad (11)$$

(7) Determine the variance set, $\{V(1), V(2), \ldots, V(\hat{n})\}$:

$$V(j) = E(\| x - W(j) \|^2 \mid x \in R(j)) \quad \text{for } j = 1, 1, \ldots, \hat{n} \quad (12)$$

(8) If $\hat{n} < n$ goto (3) else stop.

Step 4: Obtain the codebook search transform matrix (CSTM), h_{MN}. First, define the set C of all the subsets of the stochastic codebook,

$$C = \{cb(0), \ldots, cb(k), \ldots, cb(512 - S)\}$$

where $cb(k)$ consists of S contiguous codewords of the stochastic codebook (SC) beginning with the kth codeword

$$cb(k) = \{SC(k), SC(k+1), \ldots, SC(k+S-1)\}$$

Then the CSTM can be obtained according to

$$h(i, j) = \begin{cases} 2 \times \max_k Q(i, j, k) & \text{for } i \leqslant M, j \leqslant n \\ -1 & \text{for } n < j \leqslant N \end{cases} \quad (13)$$

where $\max_k$ will return k which maximises $Q(i, j, k)$, and $Q(i, j, k)$ is the frequency matrix which accumulates the frequency of $a(1)$ of the synthetic signals in the interval $(u(i-1), u(i))$. The constant 2 represents the offset factor due to the consecutive two codevectors of the stochastic codebook overlapping 58 samples. These synthetic signals are obtained from filtering all the codevectors in the subcodebook $cb(k)$ with all the linear predictors in $R(j)$. Then AC cluster i and LSP cluster j can determine the start address for searching the subcodebook.

3.4 Dynamic modification of DPSCS

An incrementally updating mechanism is used such that the search structure can be modified dynamically and increase the system performance. The self-learning struture for updating the LSP classifier and the codebook search transform matrix (CSTM) is shown in Fig. 6. The incrementally updating procedure will be described in the following two steps:

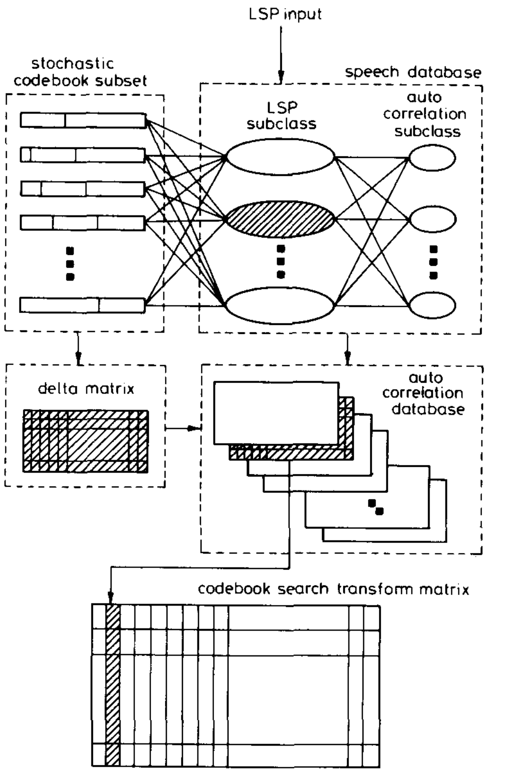

Fig. 6 *Dynamic modification structure for LSP classifier codebook search transform matrix (CSTM)*

Step 1: Determine the LSP cluster J for the input LSP parameters:

$$J = \begin{cases} \min_j d(x, W(j)) & \text{for } d(x, W(j)) \leqslant d(x, W(t)) \\ & \text{and } d(x, W(j)) \leqslant (V(t)) \\ n+1 & \text{otherwise} \end{cases} \quad (14)$$

where $1 \leqslant t, j \leqslant n$, and $\min_j$ will return j which minimises $d(X, W(j))$. If $J \leqslant n$, then we modify the $R(J)$, $V(J)$ and $W(J)$ by eqns. 9, 10 and 12; otherwise if $n < J \leqslant N$, then we reallocate and initialise $R(J)$, $W(J)$, $V(J)$ and $h(M, J)$.

Step 2: Modify the CSTM $h(i, j)$:

$$Q(i, J, k) = Q(i, J, k) + q(i, x, k) \quad \text{for all } i, k \quad (15)$$

$$h(i, j) = 2 \times \max_k Q(i, J, k) \quad (16)$$

$q(i, x, k)$ is the delta matrix which accumulates the probability of the AC of the optimal synthetic signals which lie in the interval $(u(i-1), u(i))$. The optimal synthetic signals are obtained from filtering all the codevectors in subcodebook $cb(k)$ with the input LSP parameters x.

4 Experimental results

The speech database used about 36 balanced sentences, spoken by six male and six female speakers. We used the sentences from five male and female speakers as the training sequence, and those form the remaining speakers as the test sequence. The sentences used in training were not repeated for testing. Speech signals were sampled at 8 kHz.

In the first experiment, the efficiency of the 1st-order AC classifier is studied and analysed. Fig. 4 shows the histogram of $a(1)$ of the target signal, where $a(1)$ means the value of the 1st-order AC. The cumulative distribution curves (shown in Fig. 5) between $1 - a(1)$ and the γ distribution with $\alpha = 3.664$ and $\beta = 5.722$ are very similar. The γ distribution curve can be divided into M parts with equal area, and each part indicates an $a(1)$ cluster. Because the correlation coefficient of $a(1)$ between the target and the optimal synthetic signal is 0.776, it is efficient to use $a(1)$ for classifying the synthetic signal. The relative frequency distribution is shown in Fig. 7. The same experimental conditions are also applied to the 2nd- and 3rd-order AC classifiers. The quantisation boundaries of these three-order AC classifiers are listed in Table 1, and the correlation coefficients between target and optimal synthetic signal are listed in Table 2. Finally, the high-order AC classifiers' effect on the system performance is very limited, and thus only the 1st-order AC classifier is used in our DPSCS structure to save memory space and computational effort.

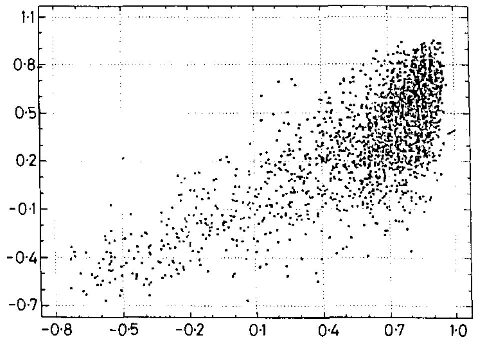

Fig. 7 *$a(1)$ experience distribution of target and optimal synthetic signal*

For the next simulation, we evaluate the DPSCS performance with various LSP cluster numbers, AC cluster

Table 1: Quantisation points of 1st–3rd-order autocorrelation coefficients with 15 clusters

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$a(1)$	−0.23	−0.47	0.08	0.17	0.24	0.30	0.36	0.41	0.46	0.51	0.56	0.60	0.65	0.71	0.79
$a(2)$	−0.13	−0.29	0.00	0.06	0.10	0.14	0.17	0.21	0.24	0.28	0.33	0.37	0.42	0.49	0.59
$a(3)$	−0.18	−0.13	−0.06	−0.02	0.00	0.04	0.08	0.11	0.14	0.17	0.21	0.24	0.29	0.35	0.45

numbers and subcodebook sizes as parameters. Both objective and subjective tests are used to measure the synthetic speech quality. On the objective tests, the influence of the LSP and AC classifiers on the hit rate and SNRSEG (segmental SNR) is estimated. Fig. 8 shows simulation results of the hit rate for the optimal codeword with various subcodebook sizes, AC cluster number as indicated and fixed LSP cluster number $N = 48$. Meanwhile, the fixed AC cluster number $M = 12$, with various LSP cluster numbers $N = 23$, 48, 57, 78, 96 and subcodebook sizes $S = 64$, 128, 256 are the parameters for evaluating the effect of the LSP classifier on the SNRSEG. The results are listed in Table 3. In general, these two experiments show that the speech quality is proportional to the cluster number of these two classifiers. On the subjective tests, the mean opinion score (MOS) [20] is used to measure the resulting speech quality. Thirty-two listeners were asked to answer a questionnaire consisting of ten questions (sentences). First, we present some samples of good (score 4) and bad (score 1) speech before the test in order to anchor the listeners. The DPSCS and static partical search coders with subcodebook sizes $S = 64$, 128, 256 and the full search coder are played randomly to the listeners. One female and one male speak each sentence on each coder twice and the listeners are required to score the synthetic speech quality. Table 4 lists the effects of the subjective tests. The comparisons of the SNRSEG and MOS between DPSCS and the static partial search are also discussed in the experiment ($N = 48$, $M = 12$). In addition to the speech quality comparisons, the extra computational effort over the static partial search is listed in the first column of Table 4. On the whole, the DPSCS can be applied to the FS1016 CELP coder directly and have good synthetic speech quality. The MOS tests show that the DPSCS with $S = 128$ has a higher score than the static partial search with $S = 256$, and it is very close to the full search quality when $S = 256$.

Table 2: Correlation coefficients between optimal synthetic speech signal and target signal

Autocorrelation	$a(1)$	$a(2)$	$a(3)$
Correlation coefficient	0.766	0.467	0.505

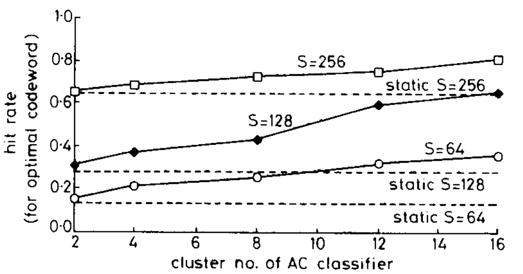

Fig. 8 *Hit rate simulation of optimal codeword with various AC cluster numbers M, subcodebook size S as indicated and fixed LSP cluster number N = 48*

Table 3: Objective tests for influence of LSP classifier on resulting speech quality in terms of SNRSEG (dB) with various subcodebook sizes $S = 64$, 128, 256, LSP clusters $N = 23$, 48, 57, 78, 96 and a fixed AC cluster $M = 12$

	S = 64, M = 12				
LSP Clusters N =	23	48	57	78	96
SNRSEG (dB)	9.2	9.3	9.3	9.5	9.7
	S = 128, M = 12				
SNRSEG (dB)	9.8	10.0	10.2	10.5	10.5
	S = 256, M = 12				
SNRSEG (dB)	10.3	10.4	10.4	10.5	10.7

Finally, a syllable /pa/ was segmented into consonant and vowel parts as test data for a typical experiment. In this case, we evaluate our system with LSP clusters $N = 48$, AC clusters $M = 12$ and subcodebook size $S = 128$. The original and synthetic speech waveforms of the consonant and vowel parts are shown in Figs. 9*a* and *b*, respectively.

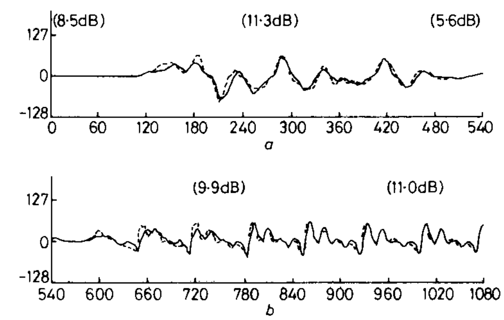

Fig. 9 *Typical experiment for (a) consonant and (b) vowel parts in speech /pa/, with LSP clusters N = 48, a(1) clusters M = 12 and subcodebook size S = 128 under our proposed DPSCS system*

——— synthetic speech waveform
– – – – original speech waveform

5 Conclusions

We have described a new dynamic partial stochastic codebook search (DPSCS) structure for the CELP speech coder. The DPSCS scheme provides a two-stage search procedure which reduces the computational effort dramatically without degrading the synthetic speech quality. From the experimental results, DPSCS always has better quality than the fixed partial search structure, at the

Table 4: Comparisons of SNRSEG and MOS between DPSCS and static partial search with subcodebook size $S = 64$, 128, 256

Codebook size	Extra comput.	SNRSEG (dB)		MOS (male)		MOS (female)		MOS (mean)	
		DPSCS	Static	DPSCS	Static	DPSCS	Static	DPSCS	Static
64	0.07	9.3	9.0	2.82	2.42	2.52	2.16	2.67	2.29
128	0.04	10.0	9.5	2.98	2.54	2.64	2.20	2.81	2.37
256	0.02	10.4	10.1	3.16	2.80	2.88	2.34	3.02	2.57
512		10.8		3.24		2.88		3.06	

small expense of extra computational complexity for selecting a potential subcodebook. The DPSCS structure has been realised with software simulation and a TMS320C30 DSP chip. From the experimental results, the performance of the modified 4.8K CELP system is faster than the conventional 4.8K CELP by about 25% without degrading the synthetic speech quality. Some further work on the DPSCS scheme will be performed, including increasing the hit rate for the optimal codeword, and on the more efficient classifier for the stochastic codebook.

6 References

1 KEMP, D., SUEDA, R., and TREMAIN, T.: 'An evaluation of 4800 bps voice coders'. Proceedings of Int. Conf. on. *Acoustics, speech, signal processing*, 1989, pp. 200–203

2 CAMPBELL, J.P. Jr., WELCH, V.C., and TREMANIN, T.E.: 'The new 4800 bps voice coding standard'. Proceedings of *Military and government speech tech.*, 1989, pp. 64–70

3 TRANCOSO, I.M., and ATAL, B.S.: 'Efficient search procedures for selecting the optimum innovation in stochastic coders', *IEEE Trans.*, 1990, **ASSP-38,** (3), pp. 385–395

4 KLEIJN, W.B., KARASINSKI, D.J., and KETCHUM, R.H.: 'Fast methods for the CELP speech coding algorithm', *IEEE Trans.*, 1990, **ASSP-38,** (8), pp. 1330–1342

5 ELSHAFEI AHMED, M., and AL-SUWAIYEL, M.I.: 'Fast methods for code search in CELP', *IEEE Trans.*, 1993, **SAP-1,** (3), pp. 315–325

6 MANO, and MORIYA: '4.8 kbit/s delayed decision CELP coder using tree coding'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, 1990, pp. 21–24

7 VASEGHI, S.V.: 'Finite state CELP for variable rate speech coding'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, 1990, pp. 37–40

8 BENYASSINE, A.: 'Mixture excitation and finite-state CELP coder'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, 1992, pp. 345–348

9 EASTON, M.G., and GOODYEAR, C.C.: 'A CELP codebook and search technique using a Hopfield net'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, 1991, pp. 685–688

10 MAKHOUL, J., SCHWARTZ, R., and EL-JAROUDI, A.: 'Classification capabilities of two-layer neural networks'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, Glasgow, Scotland, 1989, pp. 635–638

11 HARNANDEZ, L.A., CASAJUS-QUIROS, F.J., FIGUEIRAS-VIDAL, A.R., and GARCIA-GOMEZ, R.: 'On the behaviour of reduced complexity code-excited linear prediction (CELP)'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, Tokyo, Japan, 1986, pp. 469–472

12 COPPERI, M., and SERENO, D.: 'CELP coding for high quality speech at 8 kbit/s'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, Tokyo, Japan, 1986, pp. 1685–1688

13 DAVIDSON, G., and GERSHO, A.: 'Complexity reduction methods for vector excitation coding'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, Tokyo, Japan, 1986, pp. 3055–3058

14 MAUC, M., BAUDOIN, G., and JELINEX, M.: 'Reduced complexity CELP coder'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, 1994, pp. 261–264

15 KROON, P., and DEPRETTERE, E.F.: 'A class of analysis-by-synthesis predictive coders for high quality speech coding at rates between 4.8k and 16k bit/s', *IEEE J. Sel. Areas Commun.*, 1988, **6,** pp. 353–363

16 ITAKURA, F., and SUGAMURA, N.: 'LSP speech synthesizer, its principle and implementation'. Trans. of the Committee on *Speech Research*, ASJ, S79–46, November 1979

17 SOONG, F.K., and JUANG, B.W.: 'Line spectrum pair (LSP) and speech data compression'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, San Diego, CA, 1984, pp. 1.10.1–1.10.4

18 KANG, G.S., and FRANSEN, L.J.: 'Application of line spectrum pairs to low bit rate speech encoders'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, Tempa, FL, 1985, pp. 7.3.1–7.3.4

19 LINDE, Y., BUZO, A., and GRAY, R.: 'An algorithm for vector quantizer design', *IEEE Trans.*, 1980, **COM-28,** pp. 84–95

20 KITAWAKI, N., HONDA, M., and ITOH, K.: 'Speech-quality assessment methods for speech-coding systems', *IEEE Commun. Mag.*, 1984, **22,** (10), pp. 26–33

21 ATAL, B.S., and SCHROEDER, M.R.: 'Stochastic coding of speech at very low bit rates'. Proceedings of Int. Conf. *Commun.*, Amsterdam, The Netherlands, 1984, pp. 1610–1613

22 MAUC, M., and BAUDOIN, G.: 'Reduced complexity CELP coder'. Proceeding of Int. Conf. on *Acoustics, speech, signal processing*, 1992, pp. 53–56

23 DAVIDSON, G., and GERSHO, A.: 'Multiple-stage vector excitation coding of speech waveforms'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, New York, 1988, pp. 3055–3058

24 KROON, P., and ATAL, B.S.: 'Predictive coding of speech using analysis-by-synthesis techniques', *in* FURUI, S., and SONDHI, M.M. (Eds.): 'Advances in speech signal processing' (Marcel Dekker, 1991), pp. 141–164

25 KLEIJN, W.B., KRASINSKI, D.J., and KETCHUM, R.H.: 'Improved speech quality and efficient vector quantization in SELP'. Proceedings of IEEE Int. Conf. on *Acoustics, speech, signal processing*, April 1988, pp. 155–158

26 LAMBLIN, C., ADOUL, J.P., LAMBLIN, C., and MORISSETTE, S.: 'On reducing computational complexity of codebook search in CELP coder through the use of algebraic codes'. Proceedings of Int. Conf. on *Acoustics, speech, signal processing*, 1990, pp. 177–180