## ABSTRACT

The vector sum excited linear prediction coding (VSELP) can provide high quality of synthetic speech at 4.8kbps. In this paper, we propose a method to reduce computational complexity in VSELP coding. We use an overlapped sparse codebook to generate a set of VSELP basis vectors. Using these basis vectors, we can reduce the number of multiply-add operations for filtering every basis vector. We also describe an optimization method for the proposed basis vectors. Finally, we develop a 4.8kbps coder using the proposed basis vectors and perform the quality test. It has been shown that the proposed coder retains good quality of synthetic speech.


## I. INTRODUCTION

The vector sum excited linear prediction coding is one of the most efficient speech coding methods. It provides high quality of synthetic speech at 4.8kbps. The major feature of VSELP coding is that it constructs a stochastic codevector as a linear combination of the given basis vectors. Because of this feature, the complexity of the codebook search procedure can be reduced significantly compared with the original CELP coding. Besides, the codewords for the VSELP coder are more robust to channel errors.

VSELP coding is selected as the standard of North American digital cellular telephone systems. It needs a certain 40mips DSP chip, which can do floating-point multiply-add operation, to implement the full 8kbps VSELP coder for the real-time application[2]. Therefore, it is worth while reducing the complexity of VSELP coding for the low cost real-time implementation.

A lot of computational complexity in VSELP coding comes from the stochastic codebook search to decide the best excitation signal. It uses the analysis-by-synthesis method. The search procedure includes a step evaluating the zero state response of the weighted synthesis filter to every basis vector. It is remarkable that this step takes about 38% of total operations for the stochastic codebook search.

To reduce computational complexity of filtering every codevector in the stochastic codebook, the overlapped sparse codebook is used for CELP coding[1]. In this paper, we propose the overlapped sparse basis vectors generated from an overlapped sparse codebook for VSELP coding. Then, the filter responses to the proposed basis vectors can be evaluated recursively with reduced complexity. To find out the proper degree of overlapping and sparsity of proposed basis vectors, various sets of the basis vectors are tested.

The basis vectors for a VSELP codebook can be optimized over a training database[2]. Because of the overlapping property of the proposed basis vectors, we cannot train them using the algebraic methods as described in [2]. However, using steepest descent method, we also can train the overlapped sparse basis vectors and achieve the improvement in weighted segmental SNR.

## II. STOCHASTIC CODEBOOK SEARCH

VSELP coder uses the analysis-by-synthesis method to select the excitation codevectors from given codebooks. In this work, we use one stochastic codebook for the randomness of the excitation signal and one adaptive codebook for the periodicity of the excitation signal caused by the vibration of the vocal cord. As mentioned before, in VSELP coder, M basis vectors $v_m$ construct the stochastic codebook, so a stochastic codevector $u_i$ may be defined as (2.1).

$$u_i(n) = \sum_{m=1}^{M} \theta_{im} v_m(n) \,, \tag{2.1}$$

for $0 \le i \le 2^M-1$ and $0 \le n < N$

$$\theta_{im} = \begin{cases} +1 & \text{if (bit m of codeword i) = 1} \\ -1 & \text{if (bit m of codeword i) = 0} \end{cases}$$

To shape the spectrum envelope of the excitation signal, VSELP coder uses an LPC all-pole filter as the synthesis filter. The filter coefficients are estimated in every frame and updated once a subframe through interpolation, where a frame consists of four subframes. The codebook gains and indices are selected in every subframe to minimize the perceptually weighted error[2].

Because of the computational load, the VSELP coder searches the codebooks sequentially. It searches the adaptive codebook first and then searches the stochastic codebook. As a result, we must be satisfied with the suboptimal excitation signal, although the selected codevectors are jointly optimized through orthogonalization. After all the codevectors are selected, the gains are evaluated simultaneously.

Let p(n) be the target signal of codebook search procedure. We can produce p(n) by subtracting the zero input response of the bandwidth expanded synthesis filter H(z) from the weighted input speech[2]. Then, the whole codebook search procedure selects the most similar synthesized signal p'(n) with respect to the optimal codewords.

First, the adaptive codebook search procedure selects the codeword l, which minimizes the error $E'_l$ of (2.2), assuming $\beta'$ is the optimal gain.

$$E'_l = \sum_{n=0}^{N-1} \left(p(n) - \beta' b'_l(n)\right)^2 \tag{2.2}$$

where $b'_l$ is the zero state response of H(z) to an adaptive codevector $b_l$. Because of the free gain term $\beta'$, we can interpret the adaptive codebook search procedure as selecting the codevector $b_l$ which minimizes $\cos^2 \omega_{pb'_l}$, where $\omega_{pb'_l}$ is the angle between two vectors p and $b'_l$.

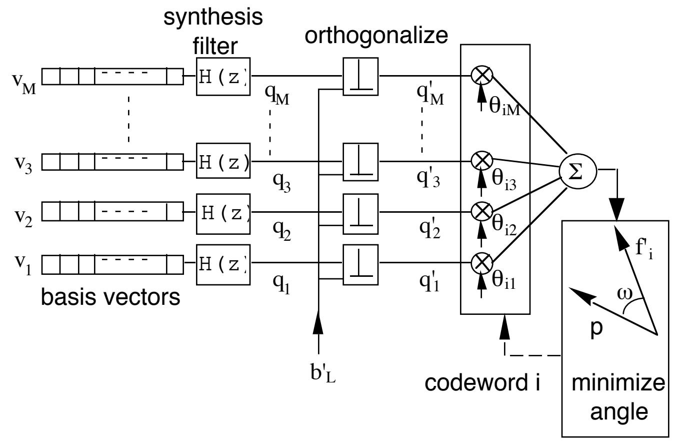



**Figure 1. Stochastic codebook search**

After finding the optimal adaptive codevector $b_L$, every weighted stochastic codevector must be orthogonalized to $b'_L$. Fortunately, for the VSELP codebook, this task is reduced to orthogonalizing only M weighted basis vectors $q_m$ to $b'_L$. Let $q'_m$ be $q_m$ after orthogonalization to $b'_L$. Then the weighted codevector $f_i$ after orthogonalization, $f'_i$, can be expressed as:

$$f'_i(n) = \sum_{m=1}^{M} \theta_{im} q'_m(n), \tag{2.3}$$

for $0 \le i \le 2^M-1$ and $0 \le n < N$

Now we can find the excitation codeword i which minimizes:

$$E'_i = \sum_{n=0}^{N-1} (p(n) - \gamma' f'_i(n))^2 \tag{2.4}$$

where $\gamma'$ is the optimal gain for each $f'_i$. As the case of the adaptive codebook search, the optimal codevector $f'_I$ minimizes $\cos^2 \omega_{pf'_i}$, where $\omega_{pf'_i}$ is the angle between two vectors p and $f'_i$. Figure 1 depicts the stochastic codebook search procedure conceptually.

## III. OVERLAPPED SPARSE BASIS VECTORS

Analyzing the amount of multiply-add operations for the stochastic codebook search, we have found that evaluating the filter responses to the basis vectors consume 38% of total operations.

It is well known that the overlapped sparse codebook reduces computational complexity of filtering every codevector greatly. The filtering results are very similar to those of the conventional codevectors if all the codevectors are random sequences.

In fact, the basis vectors of VSELP coder may be the random sequences. So if the basis vectors are N dimensions, we can generate M basis vectors $v_m$ from an overlapped codebook C as (3.1).

$$v_m(n) = C[M\alpha - m\alpha + n], \tag{3.1}$$

where the codebook C is realized by a one dimensional array, containing $N+\alpha(M-1)$ samples of random values. $\alpha$ is the number of shifts between two adjacent basis vectors. Hence, $\alpha$ decides the degree of overlapping.

Now we can evaluate the filter response $q_m$ to every basis vector $v_m$ recursively like (3.2).

$$q_1(n) = \sum_{k=0}^{n} v_1(k)\ h(n-k)\ ,\quad \text{if } 0 \le n < N \tag{3.2}$$

$$q_m(n)= \begin{cases} \displaystyle\sum_{k=0}^{n} v_m(k)\ h(n-k), & \text{if } 0 \le n < \alpha \\ \displaystyle q_{m-1}(n-\alpha)+\sum_{k=0}^{\alpha-1} v_m(k)h(n-k), & \text{otherwise} \end{cases}$$

for $1 < m \le M$

where h(n) is the zero state impulse response of H(z). If C contains zeros with probability $\sigma$, the number of multiply-add operations for evaluating all $q_m$ reduces to:

$$(1-\sigma)\left[\frac{(N+1)N}{2}+(M-1)\left\{\frac{(\alpha+1)\alpha}{2}+(N-\alpha)\alpha\right\}\right].$$

The basis vectors in VSELP coder can be optimized over a training speech data. The optimization should minimize the total weighted error. We can express the total weighted error E for each subframe as a function of the given basis vectors.

$$E = \sum_{n=0}^{N-1} |\,e(n)\,|^2 \tag{3.3}$$

$$e(n) = [p(n) - \beta b'_L(n) - \gamma\ f_I(n)] \tag{3.4}$$

$$f_I(n) = \sum_{m=1}^{M} \theta_{Im}[h(n)*v_m(n)] \tag{3.5}$$

where $\beta$ and $\gamma$ are the optimal gains, and indices I and L are the optimal codewords of stochastic and adaptive codebook, respectively.

Then, the partial derivative of E with respect to each sample of the basis vectors is estimated as:

$$d_k(n) = \frac{\partial E}{\partial v_k(n)} = -2\gamma\ \theta_{Ik} \sum_{n=0}^{N-1} e(n)h(n-k) \tag{3.6}$$

To minimize E, we can compute the optimal basis vectors by solving the simultaneous equations that result from setting all the partial derivatives equal to zero[2]. Alternatively, we also can update $v_k$ like (3.7) iteratively according to the steepest descent method[5].

$$v_k^{t+1}(n) = v_k^t(n) - \mu d_k^t(n) \tag{3.7}$$

where the parameter with superscript t means the value at time t. The constant $\mu$ is the learning rate, that are related to the stability and speed of convergence of the optimization.

In the overlapped basis vectors, one sample of the overlapped codebook stands for several samples of different basis vectors. Hence, the simultaneous equations are overdetermined and there may be no solution. However, we can update the overlapped basis vectors with d[l]. d[l] is the sum of all the partial derivatives to the samples of the basis vectors which come from the same sample of the overlapped codebook as (3.9).

$$C^{t+1}[l] = C^t[l] - \mu d^t[l] \tag{3.8}$$

$$d[l] = \sum_{k=1}^{M} d_k(l+k\alpha-M\alpha) \tag{3.9}$$

where $d_k(n) = 0$ for $n \ge N$ or $n < 0$.

After updating, the new basis vectors do not retain their sparsity. So we must center-clip those vectors to have enough zeros. During the optimization, we can get more improvement if we increase the sparsity from zero to the target value, a little at a time.

## IV. EXPERIMENTS

### 4.1 Environments

All the VSELP coders for our experiments are implemented identically except the basis vectors. Table 1 shows the bit allocations, and update timings for our VSELP coder at the 8kHz sampling rate. LPC coefficients are converted to LSP's and then quantized in every frame. The adaptive codebook search procedure follows the method for DoD 4.8kbps CELP coder: using noninteger pitch lags, delta searching, hierarchical searching and modified minimization of the weighted error[3]. As to the hierarchical searching, the optimal integer pitch lag is selected first and then nearby noninteger pitches are tested. The adaptive codebook gain and the stochastic codebook gain are quantized independently using nonuniform quantization levels made by LBG algorithm.

For training the basis vectors and the gain quantizers, we use a training data, that is about 29 seconds long. It contains the voices of two male speakers and one female speaker. The test data including the training data is about 52 seconds long and contains two more male and female speakers.

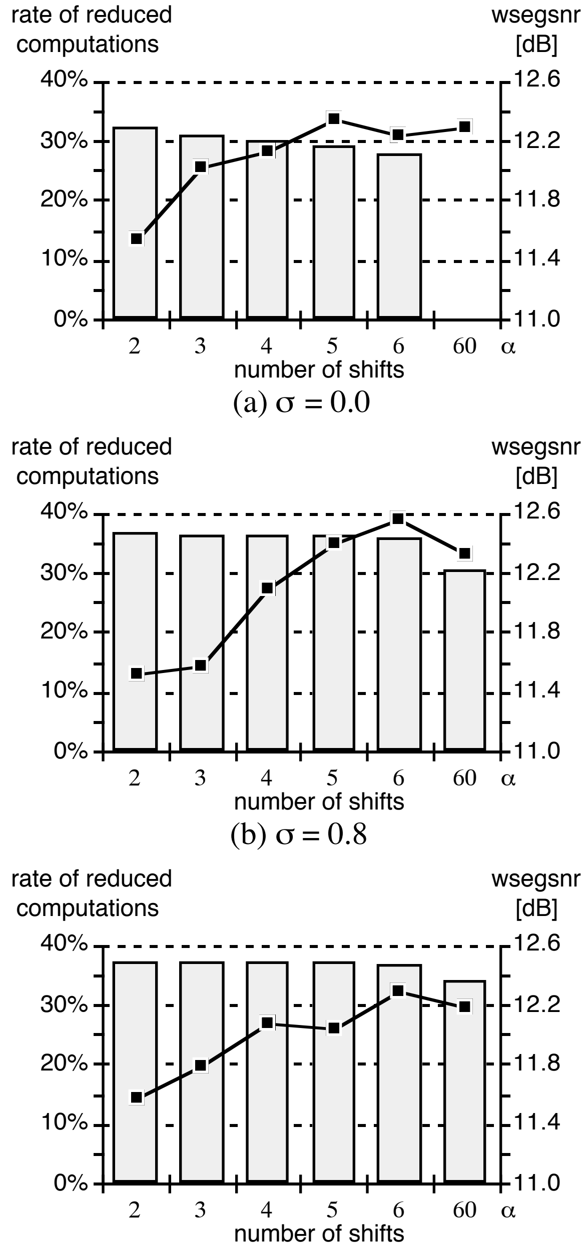
### 4.2 Results

First, we observe the degradation of speech quality caused by using the overlapped sparse basis vectors. We have tested 18 sets of the basis vectors, which have several levels of the sparsity and overlapping. We generate all these sets from the white random sequence, and make them sparse by the center-clipping. Figure 2 shows the changes in the speech quality (curves) and the percentage of the reduced computations (bars) according to the degree of overlapping. Note that the quality of the coder is retained if the number of shifts $\alpha$ is larger than 5. The sparsity $\sigma$ of the basis vectors does not affect the quality seriously until we keep it smaller than 0.9.

**Table 1. Bit allocation for 4.8kbps VSELP coder**

| | linear predictor | adaptive codebook | stochastic codebook |
|---|---|---|---|
| update | 30 ms | 7.5 ms | 7.5 ms |
| parameters | 10 LSPs | 256 codewords | 2048 codewords |
| bits per frame | 34 (3,4,4,4,4, 3,3,3,3,3) | index: 8,6,8,6<br>gain: 5x4 | index: 11x4<br>gain: 4x4 |

(Note : two bits per frame are unused.)




(a) $\sigma = 0.0$

(b) $\sigma = 0.8$

(c) $\sigma = 0.9$

**Figure 2. Results of the quality tests of overlapped sparse basis vectors. (a) sparsity 0.0 (b) sparsity 0.8 (c) sparsity 0.9**

For example, we can retain the quality of synthesized speech by using a set of basis vectors with $\alpha$ = 5 and $\sigma$ = 0.8, while 36.1% of the operations for the stochastic codebook search are reduced. By the way, we do not quantize the gains during this experiment for convenience.

When we select a set of the basis vectors with proper degree of overlapping and sparsity, we can optimize them over a training data as mentioned in section 3. At this time, we optimize the basis vectors with $\alpha = 5$ and $\sigma = 0.8$ whose performance has been shown in figure 2b. In figure 3, we plot the improvement in weighted segmental SNR for each iteration during the optimization. We update the basis vectors once per 30 frames like the batch mode training of the neural network. The learning rate $\mu$, whose initial value is 0.1, decreases exponentially for each iteration until it becomes smaller than 0.01.

Initially, the basis vectors give 12.80 dB in weighted segmental SNR for the training data. This value increases to 13.57 dB after 6 iterations.

We compare the performance of the proposed basis vectors, which are trained previously, with the conventional basis vectors. Table 2 summarizes the result of the comparison. As you can see in table 2c, the weighted segmental SNR of the proposed coder is smaller than that of the conventional coder only by 0.16 dB for the test data. During the informal listening test, we cannot discriminate the synthesized speech of the proposed coder from that of the conventional coder. It should be noticed that the quality improvements due to the optimization are also shown for speech data outside the training data.

## V. CONCLUSION

In this paper, we have described a complexity reduction method for the VSELP coding. Using the overlapped sparse basis vectors, the computational load to calculate the weighted synthesis filter response to every basis vector is significantly reduced. We have tested various sets of the basis vectors concerning the degree of overlapping and sparsity. As the result of the comparison, it has been shown that the speech quality is retained if we shift the overlapped basis vector from its adjacent vector by more than 5 samples and keep the sparsity smaller than 0.9.

We also have described the optimization method for the overlapped sparse basis vectors using the steepest descent method. After the optimization, the subjective and objective qualities have noticeable improvements. These improvements are retained for the speech data outside the training data.

We implemented a VSELP coder with reduced complexity using the overlapped sparse basis vectors. This coder needs only 64% of multiply-add operations for the original stochastic codebook search and the speech quality is scarcely distinguishable from that of the conventional coder.

## REFERENCES


[1] G. Davidson, M. Yong, A. Gersho, "Real-time vector excitation coding of speech at 4800 bps", Int. Conf. Acoust. Speech. Signal Processing, 1987, pp.2189~2192

[2] I. A. Gerson and M. A. Jasiuk, "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8kbps", Int. Conf. Acoust. Speech. Signal Processing, 1990, pp.461~464

[3] Federal Standard 1016, "Analog to Digital Conversion of Radio Voice by 4.8kbps CELP", Feb. 14, 1991

[4] M.E. Ahmed, "Fast Methods for Code Search in CELP", IEEE trans. speech and audio processing, Vol. 1, No. 3, 1993, pp.315~325

[5] J.A. Freeman, D.M. Skapura, *Neural Networks Algorithms, Applications, and Programming Techniques* , Addison-Wesley, 1991, pp.57~68

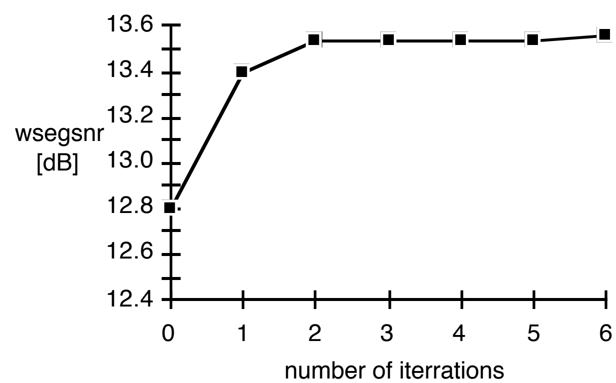




**Figure 3. Overlapped sparse basis vector optimization**

**Table 2. Coder performance measured by weighted segmental SNR. (a) initial coder (b) after optimization (c) after gain quantization**

| | conventional basis vectors with $\alpha = 60$, $\sigma = 0$ | | proposed basis vectors with $\alpha = 5$, $\sigma = 0.8$ | |
|---|---|---|---|---|
| | training data | test data | training data | test data |
| (a) | 12.67 dB | 12.30 dB | 12.80 dB | 12.41 dB |
| (b) | 13.77 dB | 13.15 dB | 13.57 dB | 12.95 dB |
| (c) | 13.38 dB | 12.77 dB | 13.21 dB | 12.61 dB |