A FAST METHOD FOR CODEBOOK SEARCH IN VSELP CODING

Sung Joo Kim, Yung Hwan Oh

Department of Computer Science, Korea Advanced Institute of Science and Technology
373-1 Kusong-dong, Yusong-gu, Taejon, 305-701, Republic of Korea

ABSTRACT The vector sum excited linear prediction(VSELP) coding gives high quality of synthetic speech at bit rates as low as 4.8kbps, but its computational complexity is prohibitive for real time applications. In this paper, we propose a method to reduce the computations of the VSELP codebook search procedure. The proposed method reduces the search space efficiently, before applying every linear combination of the basis vectors to the codebook search procedure. It decides whether it can fix the combination coefficient of each basis vector using heuristics so that the number of combinations decreases. It has been shown that the proposed method retains good quality of synthetic speech and reduces the computations of codebook search procedure by more than 40% of the origin.

1. INTRODUCTION

When we want to transmit speech signals through the digital media or to store them into the digital storage, we must code the speech signal digitally. To use the given media efficiently, many methods for speech coding are developed. The vector sum excited linear prediction coding is also one of most efficient speech coding methods. It falls into the class of the vector excitation coding the same as the code excited linear prediction(CELP) coding[1].

By using VSELP coding, we can code the speech signal at bit rate as low as 4.8kbps and the quality is good enough to adopt this method for telephony. The major feature of VSELP coding is that the excitation codevectors are constructed as a linear combination of basis vectors. Because of this feature, the computations of codebook search procedure in VSELP coding can be greatly reduced compared with the original CELP coding, and the excitation codewords for the VSELP coder are more robust to bit errors caused by channel errors. VSELP coding was selected as the standard for use in North American digital cellular telephone systems and the full 8kbps VSELP coder has been implemented using a certain 40mips DSP chip which can do floating-point multiply-add operation[2].

In this paper, we present a fast codebook search method so that the VSELP coder with reduced computational complexity can be implemented with less cost. The key idea of the proposed codebook search method is the heuristic decision over some combination coefficients of the basis vectors at the preprocessing stage. We implemented a VSELP coder using the fast codebook search method and evaluated its performance.

2. VECTOR SUM EXCITED LINEAR PREDICTION CODING

2.1 Overview of the coder

VSELP coding uses the analysis-by-synthesis method to choose the excitation codevectors from given codebooks. It has one or more stochastic codebooks for the randomness of excitation signal and one adaptive codebook to imitate the periodicity of excitation signal caused by the vibration of the vocal cord. In this study, we use only one stochastic codebook. A bandwidth expanded LPC all-pole filter is used as the synthesis filter. It shapes the spectrum envelope of the excitation signal. 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 total weighted error.

Because of the computational load, the VSELP coder searches the codebooks sequentially. First, the adaptive codebook index is chosen and then the stochastic codebooks are searched. To achieve the joint optimization during sequential searching, the weighted stochastic codevectors are orthogonalized to previously selected weighted excitation signals. After all the codevectors are selected, the gains are evaluated simultaneously.

2.2 Codebook search procedure

As mentioned above, the linear combinations of M basis vectors v_m's are used as stochastic codevectors in VSELP coding, and the combination coefficients θ_{im}'s are restricted to have either 1 or -1 according to the codeword. (2.1) defines a stochastic codevector u_i.

$$u_i(n) = \sum_{m=1}^{M} \theta_{im} v_m(n), \qquad \text{for } 0 \le i \le 2^M\text{-}1 \text{ and } 0 \le n < N \tag{2.1}$$

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

Assuming the target signal p(n) in VSELP coding, the VSELP codebook search procedure selects the indices L and i which minimize the total weighted error $E_{L,i}$ of (2.2).

$$E_{L,i} = \sum_{n=0}^{N-1} \left(p(n) - \beta b'_L(n) - \gamma f_i(n)\right)^2 \tag{2.2}$$

where b'_L and f_i are the zero state response of H(z) to an adaptive codevector b_L and a stochastic codevector u_i, respectively. As (2.2) shows, we must search all the combinations of adaptive codevectors and the stochastic codevectors to find out the optimal excitation signal, but this strategy requires too many computations. Therefore, in real implementation, we do the sequential search and must be satisfied with the suboptimal excitation signal, while the selected codevectors are jointly optimized through orthogonalization.

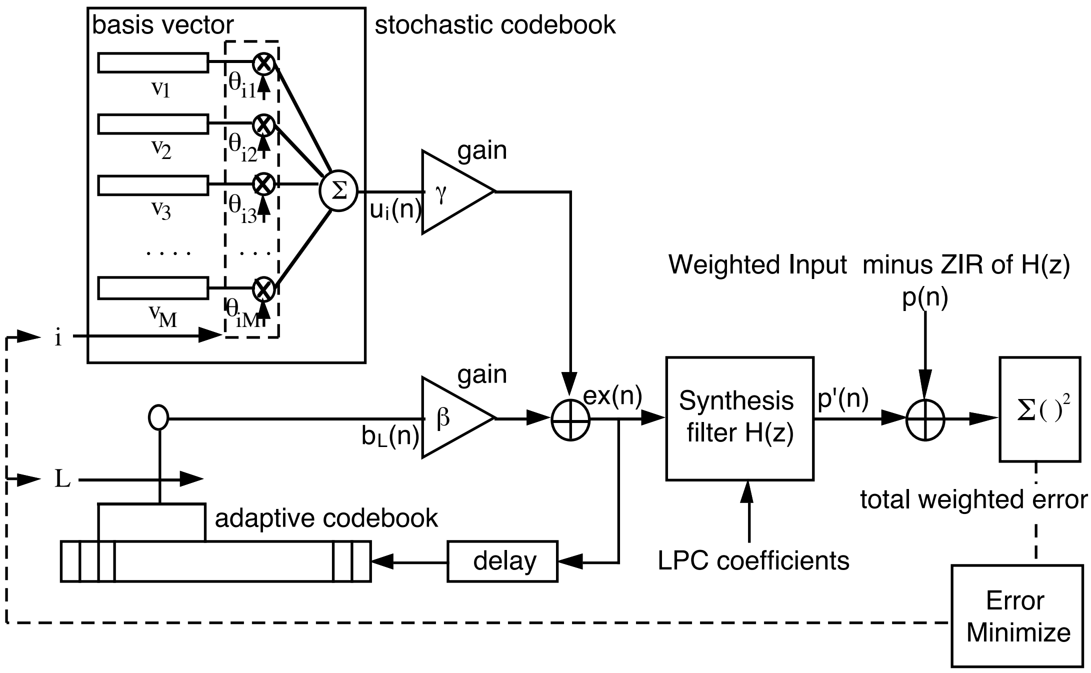


Figure 1 - VSELP coder

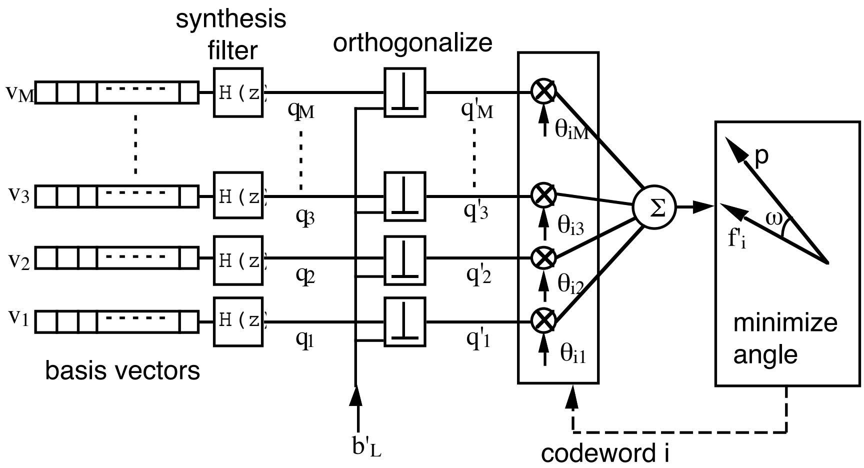


Figure 2 - Stochastic codebook search

In the adaptive codebook search procedure, the adaptive codeword L which minimize the error E'_L of (2.3) is chosen, where β' is optimal gain for each weighted codevector b'_L.

$$E'_L = \sum_{n=0}^{N-1} \left(p(n) - \beta' b'_L(n)\right)^2 \qquad (2.3)$$

Because of the free gain term β', the adaptive codebook search procedure will find a codevector b_L which minimizes the $\cos^2 \omega_{pb'_L}$, where $\omega_{pb'_L}$ is the angle between two vectors p and b'_L.

After finding the optimal adaptive codevector b_L, every weighted stochastic codevector must be orthogonalized to b'_L to achieve joint optimization. This task reduces to orthogonalizing only M weighted basis vectors q_m to b'_L for the VSELP codebook. Let q'_m be the orthogonalized weighted basis vectors, then the orthogonalized weighted codevectors f'_i can be expressed as:

$$f'_i(n) = \sum_{m=1}^{M} \theta_{im} q'_m(n) \;, \text{ for } 0 \le i \le 2^M\text{-}1 \text{ and } 0 \le n \le N\text{-}1 \qquad (2.4)$$

Now we can find the optimal excitation codeword i which minimize:

$$E'_i = \sum_{n=0}^{N-1} \left(p(n) - \gamma' f'_i(n)\right)^2 \qquad (2.5)$$

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

3. FAST CODEBOOK SEARCH PROCEDURE

The stochastic codebook search procedure decides the combination coefficients of all basis vectors, so that the optimal codeword i is found. The original method evaluates all the 2^{M-1}

angles between the target signal p and each weighted codevector f'_i. Although we can evaluate those values recursively one by one, it takes about 44.2% of total computations of the stochastic codebook search procedure.

In this section, we propose a fast codebook search method that uses heuristic preprocessing. The proposed preprocessing will find some basis vectors whose combination coefficient can be fixed heuristically. Thus, the original search procedure decides just the remaining coefficients after the preprocessing step.

3.1 Preprocessing of codebook search

(2.5) can be rewritten as (3.1) using vector notation, where p and q'_m are N dimensional vectors.

$$E'_i = \| p - \gamma' \sum_{m=1}^{M} \theta_{im} q'_m \|^2 \qquad (3.1)$$

Because γ' is the optimal gain for given codeword i, the derivative of E'_i with respect to γ' should be zero. Using this fact, the minimizing (3.1) is equivalent to maximizing the match score of (3.2):

$$\text{match}_i = \frac{\langle p, f'_i \rangle^2}{\|f'_i\|^2} = \|p\|^2 \cos^2 \omega_{pf'_i} \qquad (3.2)$$

Since $\|p\|^2$ is a constant during the codebook search procedure, the angle between p and f'_i should be minimized to maximize (3.2). Reminding that f'_i is a linear combination of q'_m's , we can regard this task as making a vector whose direction is as near as possible to that of target vector by adding or subtracting M given vectors. To do this work, we can find a heuristic method intuitively. That is setting the combination coefficient θ_{im} to the sign of $\cos\omega_{pq'_m}$. In the experiment using a sample speech of 54 seconds long, the rate of consistency between the combination coefficient θ_{im} and the sign of $\cos\omega_{pq'_m}$ was almost 85.9%. Therefore, we can say that the sign of $\cos\omega_{pq'_m}$ is a reasonable candidate for the combination coefficient θ_{im}.

However, if we set all the θ_{im} for $1 \le m \le M$ to the sign of $\cos\omega_{pq'_m}$, the quality of synthetic speech degrades too much and the weighted segmental SNR decreases by 1.2 dB. Therefore, we need to restrict the range of the appliance of heuristic decision so that this method sets θ_{im} only if q'_m satisfies a certain condition.

3.2. Condition for preprocessing

While we compare the sign of $\cos\omega_{pq'_m}$ with θ_{im}, we can find that θ_{im} always equals the sign of $\cos\omega_{pq'_m}$ if the resemblance is big enough and the resemblance can be measured by SP_m or SC_m as follows:

$$SP_m = \|q'_m\|^2 \cos^2 \omega_{pq'_m} \qquad (3.3)$$

$$SC_m = \cos^2 \omega_{pq'_m} \qquad (3.4)$$

Using this fact, we can restrict the heuristic decision over the combination coefficient θ_{im}, only if SP_m or SC_m is bigger than the threshold. The threshold can be set with expected error rate by using statistics of above comparison test. By this restriction, the error rate of heuristic codeword selection decreases to 5%, but the average number of basis vectors whose combination coefficient is set by heuristics is only 2.8 or 3.4 using SP_m or SC_m, respectively. To increase the number of the combination coefficients which are decided heuristically, we propose K-selection method as a variable thresholding method. First, the preprocessing sets K combination coefficients which have K biggest SP_m or SC_m values among M combination coefficients. Then the other M-K coefficients can be decided using original search procedure with reduced search space of 2^{M-K} combinations instead of 2^{M-1} combinations.

According to the comparison test, the error rate is less than 5% even if we fix four combination coefficients heuristically with K-selection method at the preprocessing step.

4. PERFORMANCE EVALUATION

4.1 Environments

We implemented a 4.8kbps VSELP coder with the proposed fast codebook search method and evaluated its performance comparing with the original one. Table 1 shows the bit allocations and update timings for our VSELP coder at 8kHz sampling rate. Two bits are reserved in a frame for future expansion. 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, and modified minimization of the weighted error[3]. The basis vectors for stochastic codebook can be trained[2]. We also optimized the basis vectors over a training data sequence. The adaptive codebook gain and the stochastic codebook gain are quantized independently using nonuniform quantization levels made by LBG algorithm.

We adopted the weighted segmental SNR to measure the quality of the synthetic speech and informal listening test has been performed.

4.2 Results

In stochastic codebook search procedure, about 53,175 times of the multiply-add operations are needed in every subframe and the 44.2% of those are consumed to evaluate $match_i$ of (3.2) for $1 \leq i \leq 2^{M-1}-1$. So the rate of reduced computations to the total computations for stochastic codebook search procedure as the result of heuristic decision over K combination coefficients can be calculated as:

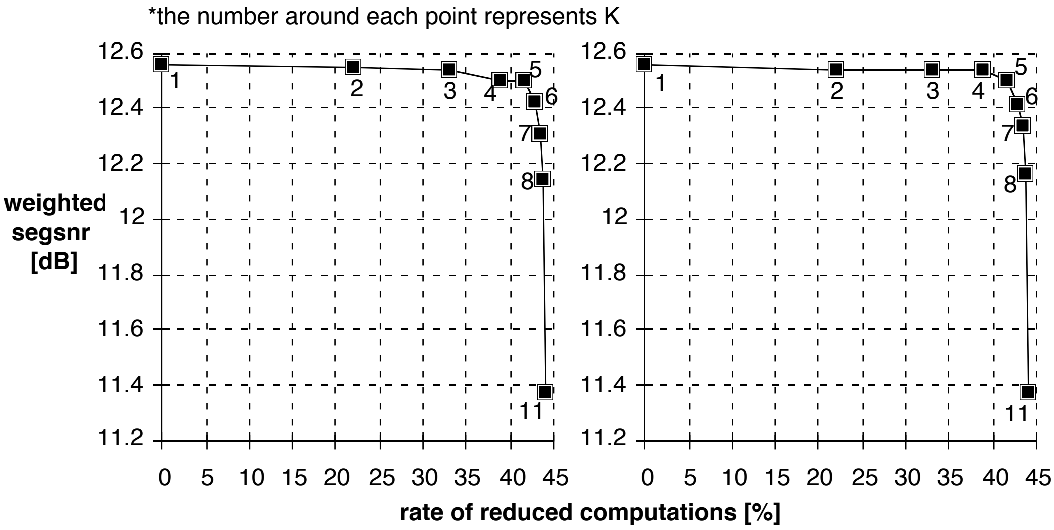
$$S_K = 44.2 \times \{1 - (2^{M-K} -1) / (2^{M-1} -1)\} \ \ [\%] \qquad (4.1)$$

while additive computations for computing SP_m or SC_m and K-selection is small enough to neglect.

	linear predictor	adaptive codebook	stochastic codebook
update	30 ms	7.5 ms	7.5 ms
parameter	10 LSP	256 codeword	2048 codeword
bits per frame	34 (3,4,4,4,4, 3,3,3,3,3)	index: 8,6,8,6 gain: 5x4	index: 11x4 gain: 4x4

(note : two bits per frame are unused.)

Table 1 - Bit allocations for 4.8kbps VSELP coder



(a) selecting K biggest SP_m (b) selecting K biggest SC_m

Figure 3 - Performance of proposed preprocessing

In the experiment, we varied the value K from 1 to 8. For each case, we calculated the rate of reduced computations and measured the speech quality by means of the weighted segmental SNR. Figure 3.a and 3.b show the results of experiment with respect to using SP_m and SC_m as the resemblance measurement.

As the result, the rate of reduced computations became 41.5%, while the weighted segmental SNR decreased only 0.06 dB, if K is set to 5 and SP_m is used. In the informal listening test, we could not find any significant difference between the synthetic speech quality of proposed and original VSELP coder.

5. CONCLUSION

In this paper, we describe a fast codebook search method for the VSELP coding. The proposed method selects some basis vectors by using a resemblance measurement and then sets their combination coefficients before the search procedure. Hence, the search space is reduced and the computations for the evaluation of the match score for all possible stochastic codewords are also reduced.

In the experiment, we tested two kinds of resemblance measurements and varied the number of basis vectors whose combination coefficients are decided by the proposed preprocessing. Using the proposed method, we can reduce the computations for stochastic codebook search procedure by more than 40% and the quality of synthesis speech are retained.

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