Enhanced Wavelet Transform-based CELP Coder with Band Selection and Selective VQ

대역 선택 구조와 선택적 벡터 양자화를 이용한 개선된 웨이브릿 변환형 CELP 부호화기

장 동 일*, 조 영 권*, 안 수 길*
(Dong-il CHANG*, Young-Kwon CHO* and Souguil ANN*)

ABSTRACT

In this paper, we present a new wavelet transform-based CELP coder, called band selection wavelet transform CELP (BS-WTCELP) operated at 4.8 kbps. The proposed algorithm uses a band selection scheme of frequency bands of wavelet transform and selective vector quantization (VQ). The band selection and selective VQ structure is implemented by using a classified VQ structure. The proposed algorithm has about 0.5-1.0 dB improvement in segmental SNR compared with the conventional CELP that uses the random codebook search, while it has significantly reduced computational and storage complexity. Many experimental results have shown that the proposed algorithm is more suitable for most real-applications than the conventional CELP and wavelet transform CELP.

요 약

본 논문에서는 대역선택 웨이브릿 변환 CELP 부호화기라 명명한 4.8 kbps 전송률의 새로운 웨이브릿 변환형 CELP 부호화기를 구현하였다. 제안된 알고리듬에서는 이산 웨이브릿 주파수 대역에 대한 대역 선택과 선택적 벡터 양자화 기법을 사용하였다. 이러한 대역 선택 및 선택적 벡터 양자화 구조는 구분형 VQ 구조를 이용하여 구현하였다. 제안한 알고리즘은 계산량 및 저장용량을 크게 줄이면서도, 기존의 불규칙 잡음 코드북 검색 구조에 비해 0.5 에서 1 dB 가량 개선된 segmental SNR을 갖는다. 많은 실험 결과를 통해 확인한 결과, 제안된 대역 선택 웨이브릿 변환 CELP 부호화기는 기존의 CELP 구조나 웨이브릿 변환 구조에 비해서 실제 응용에 훨씬 적합함을 확인하였다.

Ⅰ. INTRODUCTION

In low rate speech coding applications at rates above 4.8 kbps, the code-excited linear prediction (CELP) coder can produce good quality speech, but it has several problems in terms of complexity and performance. The CELP coder requires huge computational and storage complexity due to the random codebook search. The huge computational

*Department of Electronics Engineering,
접수일자 : 1994년 10월 29일

complexity that is the major obstacle in implementation of CELP is due to the closed-loop configuration of random codebook search. In the random codebook search, an operation of synthesis filtering should be performed on every codeword of random codebook. Also the large size of the random codebook requires huge storage complexity. Besides the complexity, CELP has poor reproduction quality in transitions from unvoiced to voiced speech where the pitch prediction shows poor performance. At the starting points of the voiced speech, the components in the memory of pitch predictor do not have enough periodicity to reproduce the periodicity in voiced speech accurately. After some periods are passed, they can reproduce it with enough precision. Because of the poor performance of pitch prediciton, the overall performance of CELP based on the random codebook model is significantly degraded in this region.

Many algorithms have been proposed to improve the performance and reduce the computational complexity of CELP [1]-[4]. Recently, Ooi *et al.* proposed a *wavelet transform-based CELP* (WT-CELP) coder which has an open loop search for the residual speech of pitch predicition to reduce the computational complexity of random codebook search [4]. In this WT-CELP, only two DWT coefficients with the highest amplitudes are selected and the average magnitude of two coefficients is quantized by a scalar quantizer. With the information about the magnitudes, the position information of the two coefficients is also encoded. This coder has significantly reduced computational and storage complexity, because it needs no filtering operation in searching procedure and has a scalar quantizer structure. In spite of the rough coding of WT coefficients, the WT-CELP coder has considerable reproduction quality compared with the conventional CELP in voiced speech. In unvoiced speech, however, it is difficult to model the residual speech by the two-peak model of the WT-CELP with enough randomness. Because the unvoiced speech reproduced by WT-CELP is too static, a hybrid structure of WT codebook with random codebook is necessary. The hybrid structure of WT-CELP adopts a U/V detection scheme to detect unvoiced regions and perform the closed-loop random codebook search instead of the open-loop WT codebook search on unvoiced speech. The need of hybrid structure causes great reduction of efficiency of the WT-CELP in computational and storge complexity.

In this paper, we propose a new WT-based CELP coder, a *band selection and selective VQ* of DWT coefficients, which may be also used in unvoiced speech. To overcome the limitations of the two-peak model of the conventional WT-CELP, the proposed coder, named by *band selection wavelet transform CELP* (BS-WTCELP), exploits a band selection and selective VQ scheme. A classified VQ structure is used for performing efficiently the band selection and selective VQ. As the results, the proposed coder with an open-loop WT excitation coding can provide high quality speech in both voiced and unvoiced speech, while reducing the computational complexity significantly. To compare the performance of the proposed algorithm with WT-CELP and CELP, we use segmental SNR of synthesized speech as a performance measure and also perform a simple preference test to compare the subjective quality. Besides the performance, we also compare the required complexity in storage and computation. Experimental results have shown that the BS-WTCELP coder shows better reproduction quality than the conventional CELP and WT-CELP, while significantly reducing the complexity compared with that of the CELP.

This paper is organized as follows. In section II, we discuss the basic principles of the DWT. In section III, we provide the details of the proposed band selection and selective VQ of DWT coefficients, and we describe the overall structures of the BS-WTCELP coder and its implement ation in section IV. Performance comparison is in section V and conclusions are followed in section VI.

II. DWT : THE BASIC PRINCIPLES

Wavelet transform (WT) is an efficient and powerful tool of signal analysis. Using a well-localized function in both time and frequency domain, called *wavelet*, the WT provides a time-frequency representaiton of a signal. A finite-scale wavelet expansion of a signal $x(n)$ can be represented by

$$x(n) = \sum_{l=L}^{\infty} \sum_{k=-\infty}^{\infty} \sqrt{2^l}\, d_l(k) h(2^l n - k)$$
$$+ \sum_{k=-\infty}^{\infty} \sqrt{2^L}\, a_L(k) g(2^L n - k), \qquad (1)$$

where $h(n)$ is a wavelet and $g(n)$ is a scaling function. In (1), $a_L(n)$ corresponds to the approximaton of $x(n)$ up to scale 2^L, and $d_l(n)$'s correspond to the details of $x(n)$. The approximation $a_L(n)$ and the details $d_l(n)$'s are called the discrete wavelet transform (DWT) of $x(n)$. In real applications, we can let $x(n)$ be $a_K(n)$ since we usually deal with a finite samples of length $N=2^K$.

The wavelet $h(n)$ and the scaling function $g(n)$ are defined in terms of coefficients c_k which must satisfy several conditions such as normalization, orthogonality and regularity condition. The wavelet and the scaling function are given by

$$g(n) = \sum_{k=0}^{N-1} c_k\, g(2n-k),$$
$$h(n) = \sum_{k=0}^{N-1} (-1)^k c_{1-k}\, g(2n-k). \qquad (2)$$

The structure of DWT may be considered as a QMF filter bank. The filter bank structure of DWT is illustrated by Fig. 1. An approximation and detail of the l-1-th scale may be computed from the approximation of the l-th scale, respectively, by a recursive manner [5],

$$\mathbf{a}_{l-1} = \mathbf{L}\mathbf{a}_l,$$
$$\mathbf{d}_{l-1} = \mathbf{H}\mathbf{a}_l,$$
$$\mathbf{a}_l = \mathbf{L}^*\mathbf{a}_{l-1} + \mathbf{H}^*\mathbf{d}_{l-1}, \qquad (3)$$

where $\mathbf{a}_l = \left[a_l(0),\cdots,a_l(2^{K-l}-1)\right]^T$, $\mathbf{d}_l = \left[d_l(0),\cdots,d_l(2^{K-l}-1)\right]^T$, $\mathbf{L} = \left[c_{2i-j}\right]_{\frac{N}{2}\times N}$, $\mathbf{H} = \left[(-1)^{j+1}\, c_{j+1-2i}\right]_{\frac{N}{2}\times N}$.

Two matrices $\mathbf{L}^*$ and $\mathbf{H}^*$ correspond to the conjugate-transpose matrices of the $\mathbf{L}$ and $\mathbf{H}$, respectively. And the input $x(n)$ is the initial vector $\mathbf{a}_K$, and the constant N is the number of samples in a frame.

As the wavelet function $h(n)$ is band-pass filter $g(n)$ is a low-pass filter, $\mathbf{a}_{l-1}$ corresponds to the low-pass filtered version of $\mathbf{a}_l$, and $\mathbf{d}_{l-1}$ corresponds to the band-pass filtered version. This means that the details and approximation may be considered as the components of different frequency *bands*. In the following sections, we will treat $\mathbf{a}_l$ and $\mathbf{d}_l$'s as speech components of different DWT frequency bands. If we use a DWT filter bank of L stage, we may have L bands of details and one band of approximation $\{\mathbf{d}_{K-1}, \mathbf{d}_{K-2},\cdots,\mathbf{d}_{K-L}, \mathbf{a}_{K-L}\}$. The approximation $\mathbf{a}_{K-L}$ corresponds to the lowest frequency band and $\mathbf{d}_{K-1}$ corresponds to the highest frequency band.

In our experiments, we use the orthonormal compactly supported wavelet of Daubechies which has a length of 20 samples in DWT analysis [6], and the number of stage L is set to 4.

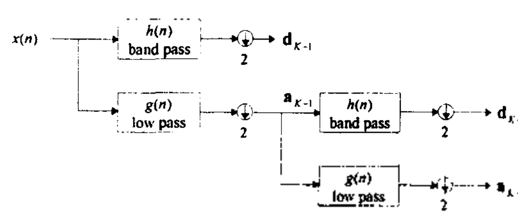

Fig. 1. QMF filter bank structure of DWT.

Ⅲ. BAND SELECTION CODING OF DWT COEFFICIENTS

In the band selection coding of DWT coefficients of the residual speech, we select several DWT frequency bands which contain most energy of the overall DWT coefficients. As we mentioned above, the performance of the conventional *coefficient selection* method is significantly degraded, when it is difficult to model the residual speech by a waveform with two peaks such as unvoiced speech. The proposed *band selection* method can avoid this inefficiency in unvoiced speech while also improving the performance in vocied speech. In fricatives and some unvoiced speech, $\mathbf{d}_{K-1}$ has the largest value of energy and the magnitudes of components are uniformly distributed. In the conventional WT-CELP, two coefficients are only selected and used in constructing the synthesized residual speech. Thus, the resulting synthesized speech has poor reproduction quality because of the insufficient reproduction of residual speech.

In the band selection algorithm, all coefficients in $\mathbf{d}_{K-1}$ should be encoded with proper precision in fricatives and some unvoiced speech. On the contrary, lower frequency bands should be also encoded accurately in vocied speech, because they have relatively larger energy than the other bands. Since they also have significantly smaller dimensions than the overall dimension N, we can reproduce them with proper precision. To select bands, we compute the magnitude of the energy ratio of each band over the total energy. The energy ratio R_l is given by

$$R_l = \frac{E_l}{\sum_{k=1}^{L+1} E_k}, \quad l = 1, \cdots, L+1, \tag{4}$$

where E_l is energy of the l-th band. We select bands starting from the largest one of the magnitude of energy ratio. For better performance of band selection, we should also consider the quantization performance of the quantizer for each selected band, in addition to the magnitude of energy of it. If all bands have nearly equal energy, we should select the bands in the order of quantization efficiency.

If we use a band selection and selective VQ, we should transmit some bits representing the status of band selection and selective VQ. The more bits we allocate to represent the status, the fewer bits may use in VQ of DWT coefficients. To make a reasonable choice of this trade-off, we exploit a classified VQ for simplicity and efficiency. In classified VQ, the input vector is classified into a class and quantized by a predetermined way for the class [7]. Similarly, in band selection coding, we classify a vector that is composed of DWT bands into a class among P classes. A predetermined way of band selection and selective VQ for the class is followed. The structure of band selection coding exploiting a classified VQ is illustrated in Fig. 2.

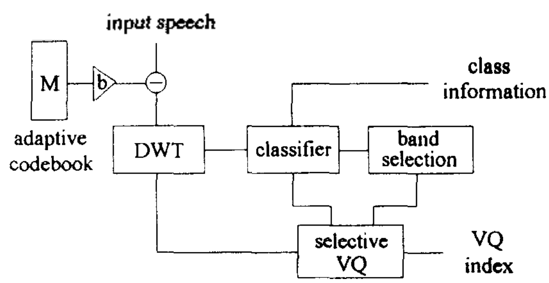

Fig. 2. The structure of band selection coding of DWT coefficients.

In Fig. 2, M and b represent the optimal pitch delay and gain, respectively, which are obtained in pitch prediction. The classifier in Fig. 2 may be a phonetic classifier, a simple VQ or else. In this paper, however, we use a simple VQ of P codewords constructed by the VQ of vectors of energy ratios. Therefore, a codeword representing a class stands for the average energy ratio of the DWT coefficient vectors in the class. This relationship may play an important role in determining a band selection for the class.

In addition to the relative energy, we should also consider the coding efficiency and characteristics of each band to achieve the best reproduction quality for determining the quantization scheme. Since we can only use some finite bits in encoding the DWT coefficients, we cannot choose and quantize whole DWT bands while maintaining the required performance. The dimensions of the upper stage bands relatively smaller than those of the lower stage bands. However, the dynamic ranges of upper stage ones are significantly larger than those of lower stage ones. The differences in dimensions and dynamic ranges should be also considered to determine the selective VQ scheme. From many related experiments, we have found some meaningful results. Firstly, the large quantization error of an upper stage detail or approximation may introduce a severe degradation of reproduction quality. Secondly, the lowest stage detail $\mathbf{d}_{K-1}$ may be quantized efficiently by a gain-shape VQ, since the quantization error in shape of it has relatively small effects on the perceptual quality compared with that in gain. Considering the characteristics of each band and experimental results, we construct an efficient BS-WTCELP coder. The proposed coder select two most significant bands and allocate the allowed bits to the selected bands considering the quantization efficiency and energy. The details of the coder structure are described in the next section.

Ⅳ. BAND SELECTION WAVELET TRANSFORM CELP CODER

The propsed band selection WT-CELP (BS-WTCELP) coder is based on the FS-1016 CELP coder and the conventional WT-CELP coder [4][8]. Input speech is sampled at 8kHz and low-pass filtered at 3.4 kHz. LPC analysis of 10-th order is performed by autocorrelation method once every 32 ms frame. LPC coefficients are converted to LSP parameters and quantized at 34 bits/frame. Pitch prediction of BS-WTCELP is performed by using only 128 integer pitch delays in odd subframe whereas 256 non-integer and integer pitch delays are used in FS-1016 CELP and WT-CELP. Delta search with only 32 integer delays is performaed in even subframe. The pitch gain is quantized by a 5 bit scalar quantizer. The reduced one bit obtained by omitting the non-integer pitch delay in each subrame is used for quantizing WT coefficients. As a result, we allocate 18 bits to the excitation coding. The bit allocation results are given in Table 1. The total bit rate of the BS-WTCELP is 4.6575 kbps without any forward error corection bits and synchronous bits. This rate is the same as that of the conventional WT-CELP.

Table 1. Bit allocation results of the proposed coder.

parameter	bits/subframe				bits/frame
	1	2	3	4	
LSP parameters	34				34
pitch lag	7	5	7	5	24
pitch gain	5	5	5	5	20
DWT coding	18	18	18	18	72
total bits/frame					150

In this paper, we use a wavelet analysis structure of 4 stages to get 4 details and an approximation every 8 ms subframe. The stage number of 4 is sufficient for our application because dividing the baseband more finely is nearly useless for enhancing the performance. We use the number of classes of 1, 2, and 4, respectively. We have found that increasing the number of classes generally causes the reduction of segmental SNR which is too small to be perceived, when the number of class is equal to or smaller than 4. We have also found that the subjective quality is slightly but gradually improved as we increase the number of class, despite of the reduction of SNR. In all the cases, we select only two bands using the modified codewords of classifying VQ. We modify the codewords composed of the energy ratios for each class considering the coding efficiency and the selection scheme of different classes. To achieve an efficient selection scheme, we make several restrictions of band selection based on the percseptual characteristics. Firstly, we only select the lower stage details when they have significantly large energy than the highest detail and the approximation. Secondly, the second detail $\mathbf{d}_{K-2}$ is excluded in selection scheme because it may be neglected without any audible distortion in most cases. Considering these restrictions, we choose the band selection scheme and bit allocation results for the number of classes of 1, 2 and 4 which are presented in Table 2.

Table 2. The results of band selection and bit allocation for the different number of classes.

Number of class	Band selection Info.	$\mathbf{a}_{K-1}$	$\mathbf{d}_{K-4}$	$\mathbf{a}_{K-3}$	$\mathbf{a}_{K-1}$
1	0	9	9	0	0
2(class1)	1	9	8	0	0
2(class2)	1	7	0	0	10
4(class1)	2	9	7	0	0
4(class2)	2	7	0	0	9
4(class3)	2	7	0	9	0
4(class4)	2	0	7	9	0

In case of only one class, $\mathbf{a}_{K-4}$ and $\mathbf{d}_{K-4}$ are selected quantized at equal rate. In voiced speech and some unvoiced speech, this single case of band selection and VQ scheme shows fairly good performance. However, in most cases of unvoiced speech, this scheme hardly reproduces the high frequency components which have the most energy. This inefficiency causes significant distortion in unvoiced speech, especially in fricatives and plosives and is very similar to that of the conventioanl WT-CELP with the structure of selecting two coefficients. Thus, this scheme is inadequate to real applications though it shows good segmental SNR in voiced speech. In order to solve this problem, we made a class to encode the unvoiced speech in which the lowest detail has relatively large magnitude of energy in case of 2 classes. To encode unvoiced speech with significant $\mathbf{d}_{K-1}$, we select $\mathbf{a}_{K-4}$ which is the approximation of waveform besides of $\mathbf{d}_{K-1}$, which compose a combination of approximation and the highest frequency detail. This scheme causes slight degradation in segmental SNR, but the perceptual quality is considerably improved since the reproduction quality of unvoiced speech is significantly improved. The reduction of SNR may result from the reduction of bits allocated to $\mathbf{d}_{K-1}$ and the misclassification in class 1. Also in class 2, the coding efficiency of $\mathbf{a}_{K-4}$ and $\mathbf{d}_{K-1}$ is not encode the WT coefficients accurately, though the reproduced speech shows reasonable subject quality. Since a little variation of the high frequency components causes hardly perceivable, the first detail $\mathbf{d}_{K-1}$ can be efficiently quantized using a gain-shape VQ structure. In spite of the sub-optimality of gain-shape VQ, the perceptual quality resulting from this gain-shape VQ is fairly well in addition to the simple structure of it.

Though the BS-WTCELP with 2 classes shows fairly well performance, we need more subdivisions to quantize the residual speech more efficiently and improve the perceptual quality of speech. In BS-WTCELP with 4 classes, we subdivide the second class of the 2 classes case into 3 classes as shown in Table 2. By using these subdivisions, we may encode some unvoiced speech and voiced speech more efficiently than the case of 2 classes. The segmental SNR obtained by the BS-WTCELP with 4 classes is slightly more decreased than the BS-WTCELP with 2 classes. On the contrary, the perceptual quality is slightly improved in some unvoiced speech. The results of performance comparison are provided in the following section in detail.

In the case of 2 classes, we can approximately select the subframe where the frequency band corresponding to $\mathbf{d}_{K-1}$ is dominant by using the estimated energy ratios computed from the LPC spectrum, instead of selecting $\mathbf{d}_{K-1}$ from the energy ratio of residual speech. When the estimated energy of it is larger than a time-varying threshold which is determined by the energy of the other bands, we only choose the band $\mathbf{d}_{K-1}$. In spite of the worse precision due to the estimation, the saved one bit obtained by omitting the class information can be used for enhancing the quantization performance. In the case of 4 classes, this estimation scheme may be also used. However, we do not use this estimation scheme in this paper, since the improvement in performance of it is hardly perceiveable.

V. PERFORMANCE COMPARISON

In this section, we compare the speech quality and coder complexity of BS-WTCELP with conventional CELP and WT-CELP. We use a FS-1016 CELP coder, the conventional WT-CELP without random codebook search in unvoiced speech, and the proposed BS-WTCELP with 2, 4 classes in performance comparison. In performance comparison, we compare the segmental SNR to represent speech quality of each coding algorithm, required storage and computational complexity. From the results of quality comparison, we could find that the BS-WTCELP coder can reproduce both voiced speech and unvoiced speech successfully, whereas the conventioanl WT-CELP cannot reproduce unvoiced speech with enough randomness and CELP cannot reproduce the rapid increasing in transition. Only in some unvoiced and steady-state voiced speech, CELP shows slightly better reproduction quality than the proposed BS-WTCELP coders with 2 and 4 classes. The comparison of synthesized speech waveforms with original speech waveform, an English word "ship", by three coders are given by Fig. 3. The waveform by BS-WTCELP is reproduced by using 2 classes.

As can be shown in Fig. 3, the conventional WT-CELP has a great degradation of performance in unvoiced speech and CELP cannot reproduce the transition with enough precision, while the BS-WTCELP reproduce unvoiced speech more exactly than WT-CELP and speech corresponding to transition more exactly than CELP.

To compare the performance of coders numerically, we construct two test data which have about one minute length. One consists of speech segments recorded by 5 male speakers and the other consists of speech segments recorded by 4 female speakers. Using these two test data, we compute the segmental SNR of 4 coders, CELP, WT-CELP, BS-WTCELP with 2 classes and BS-WTCELP with 4 classes. The values of segmental SNR are given in Table 3.

We have found that the overall segmental SNR of BS-WTCELP is higher than those of the remaining two coders by 0.5-1.2 dB in both cases with 2 and 4 classes. The increase of 1 dB in segmental SNR is considerably meaningful in low rate speech coding applications. From the results of subjective quality test by a preference test, we could easily find that the speech quality synthesized by the BS-WTCELP is clearly better than that by the conventional WT-CELP, and it is nearly equal to or slightly better than that by the CELP. These results could be easily predicted from the results of segmental SNR. From these results, we have found that the proposed BS-WTCELP coder can reproduce both voiced speech and unvoiced speech with a quality equal to or slightly better than the quality provided by the CELP. This means that the proposed WT-based coder can be used in real applications by itself, whereas the conventional WT-CELP should have a hybrid structure with CELP.

Table 3. Segmental SNR of four coders.

	CELP	WT-CELP	BS-WTCELP (2 classes)	BS-WTCELP (4 classes)
Male	8.41 dB	7.49 dB	8.53 dB	8.50 dB
Female	8.19 dB	7.83 dB	9.07 dB	8.91 dB
Overall	8.30 dB	7.66 dB	8.80 dB	8.71 dB

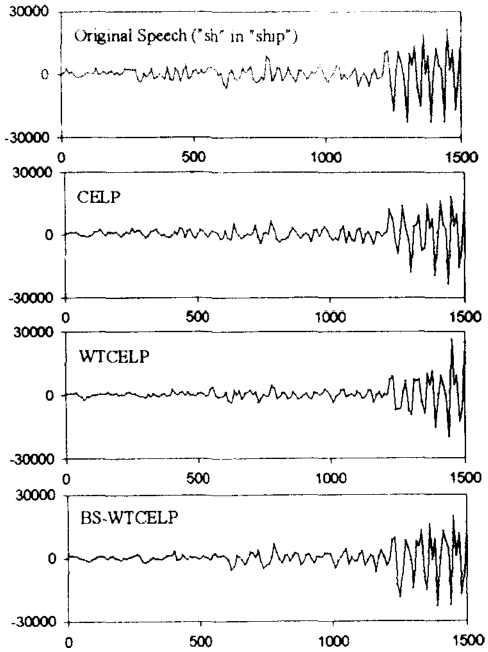

Fig. 3. Comparison of synthesized speech with original speech.

In addition to the quality comparison, we also perform the comparison of required storage and computational complexity briefly. In CELP, random codebook search requires storage for 512 codewords of 60 dimension and 32 gain codewords, which is also required in WT-CELP for encoding unvoiced speech efficiently. To reducing the required storage, a ternary overlapped codebook has been proposed and used in FS-1016 CELP coder. This ternary codebook can significantly reduce the required storage compared with the random codebook, although it causes slight degradation of performance. In BS-WTCELP of 4 classes, we only need the storage of 20 wavelet coefficients, two thousands and more codewords of dimension 4, 8, 32. However, this storage is significantly less than that of random codebook, because the number of codewords with large dimension is much smaller than that of codewords with small dimension. When CELP uses the ternary codebook instead of the random codebook, the codewords constructed by only -1, 0 and 1 that can be represented by 2 bits. Therefore, the required storage of the CELP is significantly reduced. In this case, the required storage of the proposed algorithm may be larger than that of CELP. However, the required storage of the proposed coder may not cause problems in real implementation, because it is significantly reduced value compared with that of the random codebook.

In real implementation of CELP, the major problem is huge computational complexity of it. Most examples of implementation of FS-1016 CELP have several additional structures such as the restriction of the codebook searching range which results in reduction of complexity and degradation of performance, simultaneously. The estimated computational complexities of excitation coding in CELP and the WT-CELP are about 8.3 MIPS and 1.2 MIPS, respectively [4][8]. The estimated computational complexity of excitation coding in BS-WTCELP with 4 classes is about 1.6 MIPS. It is a significantly reduced value compared with that of CELP, and a slightly increased value compared with that of WT-CELP. However, the BS-WTCELP performs pitch search only for integer pitch delays, which reduces the required computational complexity in pitch prediction. In addition, the conventional WT-CELP requires the structure of random codebook search of CELP additionally to encode the unvoiced speech efficiently. These results have shown that, in terms of the real implementation, the proposed BS-WTCELP is has the least computational complexity which is significantly reduced value compared with those of the CELP and WT-CELP.

From the above discussion, we have found that the proposed BS-WTCELP shows the best quality and the least computational complexity compared with those of the conventional CELP and WT-CELP. Though it has some increased storage requried in implementation, the BS-WTCELP is the most suitable algorithm for many applications that require lower costs and power consumption with high quality speech, such as digital mobile communication.

Ⅵ. CONCLUSION

In this paper, we have presented a band selection coding of wavelet transform excitation. The proposed coder exploits a band selection and selective VQ that is implemented by a classified VQ structure. Instead of selecting a few coefficients and quantizing them with the position information, we select several bands according to the magnitude of energy and quantize the bands by independent VQ's. To improve the perceptual performance of quantization, we consider the characteristics of DWT bands in band selection and selective VQ. The proposed band selection wavelet transform CELP can reproduce original speech successfully in both unvoiced and voiced region, whereas the conventional WT-CELP should use the random codebook search to reproduce the unvoiced speech efficiently. In addition, the open-loop configuration of the proposed coding algorithm requires significantly reduced computational complexity. In spite of the small increase in requried storage, the proposed coder is the most suitable algorithm for real application compared with FS-1016 CELP and conventional WT-CELP, since it has significantly reduced computational complexity, reasonable required storage and the highest reproduction quality. Especially, the significantly reduced complexity gives many advantages for realizing it in terms of costs and hardware complexity.

REFERENCES

1. W. B. Kleijn, "Encoding Speech Using Prototype Waveforms," *IEEE Trans. on Speech and Audio Processing*, vol. 1, pp. 386-399, 1993.
2. D. W. Griffin and J. S. Lim, "Multiband Excitation Vocoder," *IEEE Trans. on Acoust. Speech Sign. Process*, vol. ASSP-36, pp. 1223-1235, 1988.
3. R. J. McAulay, T. Parks, T. Quatieri and M. Sabin, "Sine-wave Amplitude Coding at Low Data Rates," in *Advances in Speech Coding*, B.S. Atal, V. Cuperman and A. Gersho, eds., pp. 203-214, Kluwer Academic Press, 1991.
4. J. Ooi and V. Viswanathan, "A Computationally Efficient Wavelet Transform CELP Coder," *Proc. ICASSP-94*, pp. 101-104, 1994.
5. D. Sinha and A.H. Tewfik, "Synthesis/Coding of Audio Signals Using Optimized Wavelets," *Proc ICASSP-92*, pp. 113-116, 1992.
6. I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Commun. Pure Appl. Math.*, pp. 906-996, 1988.
7. A. Gersho and R.M. Gray, *Vector quantization and Signal Compression*, Kluwer Academic Press, 1992.
8. J.P. Campbell Jr., T.E. Tremain, and V.C. Welch, "The DoD 4.8 KBPS Standard(Proposed Federal Standard 1016)," in *Advances in Speech Coding*, B.S. Atal, V. Cuperman and A. Gersho, eds., pp. 121-133, Kluwer Academic Press, 1991.

▲장 동 일

1987년 3월 : 서울대학교 전자공학과 입학
1991년 2월 : 서울대학교 전자공학과 졸업
1991년 3월 : 서울대학교 대학원 전자공학과 입학
1993년 2월 : 서울대학교 대학원 전자공학과 졸업 (공학석사)
1993년 3월 : 서울대학교 대학원 전자공학과 입학
1995년 2월 : 서울대학교 대학원 전자공학과 수료

▲조 영 권

1989년 3월 : 고려대학교 전자공학과 입학
1993년 2월 : 고려대학교 전자공학과 졸업
1993년 3월 : 서울대학교 대학원 전자공학과 입학
1995년 2월 : 서울대학교 대학원 전자공학과 졸업 (공학석사)
1995년 3월 : 서울대학교 대학원 전자공학과 입학
현재 : 서울대학교 대학원 전자공학과 재학중(박사과정)

▲Souguil Ann : Vol.11, No.1E