An Efficient Vocoder for Digital Cellular System

Sangwon Kang*, Insung Lee*, Kyongho Han* *Regular Members*

디지틀 셀룰라 시스템을 위한 효율적인 음성부호화 기술

正會員 姜 尙 遠* 正會員 李 寅 誠* 正會員 韓 敬 浩*

ABSTRACT

In this paper, we present code-excited linear prediction coding using trellis coded vector excitation, termed trellis code-excited linear prediction coding (TCELP), for an efficient 8 kbps digital cellular vocoder. A training sequence-based algorithm is developed for designing an optimized codebook subject to the TCELP structure. Also, we discuss the encoding complexity of the TCELP system and trellis symbol release rules that avoid excessive encoding delay. Finally, simulation results results for the TCELP coder are given at the bit rate of 8 kbps.

요 약

본 논문에서는 디지틀 셀룰라 시스템을 위한 효율적인 8 Kbps 음성부호화기로서, Code-excited linear prediction (CELP) 구조에 Trellis Coded Vector Excitation을 이용한 trellis code-excited linear prediction (TCELP) 음성 부호화 방식을 제시하였다. TCELP 구조에 근거한 최적 코드북 설계를 위해 traing seguence-based 알고리즘과 부호화 지연을 줄이는 trellis 심볼 release 방식이 개발되었고, 제시된 TCELP 시스템의 부호화 복잡도가 분석되었다. 끝으로 8 Kbps TCELP 부호화기의 성능이 SNR/SEGSNR과 비공식 청취시험을 통해 평가되었다.

Ⅰ. INTRODUCTION

With the popularity of the analog celluar telephone, the current analog system is straining to support current market demands. To accomodate this high demand, the analog cellular phone is rapidly transitioning toward the digital cellular phone system to increase more channel capacity, since digital transmission uses more efficient bandwidth utilization techniques. For the efficient channel utilization, demands for good quality speech coding at low bit rates have been rap-

*韓國電子通信硏究所
ETRI
論文番號 : 93-136

idly increasing in the area of digital mobile communication systems. Code-excited linear prediction (CELP) type coders [1] are considered to be very efficient low bit speech coding techniques. CELP coding does not require a scalar quantization procedure, but chooses the exitation sequence from a given codebook. Hence, two important research issures in CELP coding are the design and search procedures of a codebook. Trellis coded quantization (TCQ) [2] is a of trellis coding that labels the trellis branches with subsets of reproduction symbols. Trellis coded vector quantization (TCVQ) [3] is a generalization of TCQ that allows vector codebook branch labels. Hence the novel feature of TCVQ is the partitioning of an expanded set of vector quantization codewords into subsets and the labeling of the trellis branches with these subsets. For excitation sequences, the TCVQ encoding scheme offers a significant improvement over the vector quantization encoding scheme. In this paper, we develop an effective low-rate speech coder that incorporates TCVQ in the CELP structure. The resulting system is referred to as a trellis code-excited linear prediction coding (TCELP) system. The encoding rate under consideration is 8 kbps.

The paper is organized as follows. The TCVQ structure is discussed in Section II. Section III presents the trellis excitation encoding system. In Section IV, a training sequence-based algorithm is presented for designing an optimized codebook subject to the TCELP structure. Section V describes trellis symbol release rules that avoid an excessive encoding delay. Encoding the side information is covered in Section VI. Section VII evaluates the encoding complexity of the TCELP system. In Section VIII, simulation results for the TCELP coder are given for encoding rates of 8 kbps. A summary is provided in Section IX.

II. TCVQ STRUCTURE

Trellis coded vector quantization uses vector codewords and allows encoding at nonin teger bit rates. In [3], three different structures were given for incorporating vector quan tization with TCQ. We consider only the structure 1 formulation in this paper. Consider encoding L-dimensional source vectors at an encoding rate of R_e bits/dimension. If we assume the product $R_e L$ to be an integer, a traditional VQ would have $2^{R_e L}$ codewords. The TCVQ encoder uses a "super" codebook, S, of $2^{(R_e+\tilde{R})L}$ vector codewords, where $\tilde{R}$ is called the "codebook expansion factor"(in bits per dimension). The trellis has N states with 2^{M} branches entering and leaving each trellis state, with M an integer satisfying $M \leq R_e L$. Let $K=\tilde{R}L+M$ and partition the codebook into 2^{K} subsets, denoted as S_1, $S_1, \ldots, S_{2^K}$ each subset containing exactly $2^{R_e L-M}$ codewords. Each branch of the N-state trellis is labeled with one of the subsets. In order to assign each subset to at least one branch, there must be $N \geq 2^{\tilde{R}L}$ states.

Given the above structure and an initial state in the trellis, the encoding is performed as follows [3]:

1. For each source vector, x, find the optimal codeword and the corresponding distor tion, d_i, in each subset S_i.

2. Let the branch metric for a branch labeled with subset S_i be the distortion found in 1, and use the Viterbi algorithm [4] to determine the minimum distortion survivor path through the trellis.

M bits/vector are used to specify the best trellis path. The remaining $(R_e L-M)$ bits/vector are used to specify the codeword in the selected (branch) subset. Thus, $R_e L$ bits/vector are used to specify the sequence of codewords corresponding to the minimum distortion path through the trellis. Hence, the transmission rate is R_e bits/dimension.

The trellis design, subset construction, and branch labeling presented in [3] are summarized as follows. Ideally, the codewords in a subset should have approximately the same probability (since otherwise the subset entropy is less than the bit rate). The union of subsets corresponding to branches emanating from or entering a trellis state should be a reasonable quantzer for the source. The subsets are formed based on Ungerboeck's method of set partitioning [9], so that the Euclidean distance between the codewords in a subset is increasing with each level of partitioning. In this paper, we use Ungerboeck's trelis coded modulation trellises [9]. Consider labeling an $N-$state trellis with the 2^K sub sets partitioned from S, where K is $\tilde{R}L+M$. Denote the subsets S_i, $i=0, 1, \ldots, 2^K-1$, such that S_j and $S_{j+2^{K-1}}$, $j=0,\ldots,2^{K-1}$ are partitions of the same intermediate subset in the last partitioning step. The branches of each $N-$state trellis are then labled such that S_j and $S_{j+2^{K-1}}$ are tne subsets related to the branches entering a given trellis state. Table 1 lists example of an 8-state trellis populated with 4,8, and 16 subsets.

Table 1. Trellis Structure and Branch Labelling for an 8-state Trellis Populated with 4, 8, and 16 Subsets

		Current state							
		0	1	2	3	4	5	6	7
4 Subsets	Previous States	0	2	4	6	3	1	7	5
		4	6	0	2	7	5	3	1
	Associated Subsets	0	0	0	0	1	1	1	1
		2	2	2	2	3	3	3	3
8 Subsets	Previous States	0	2	4	6	3	1	7	5
		4	6	0	2	7	5	3	1
	Associated Subsets	0	2	0	2	1	3	1	3
		4	6	4	6	5	7	5	7
10 Subsets	Previous States	0	2	4	6	3	1	7	5
		4	6	0	2	7	5	3	1
	Associated Subsets	0	2	4	6	1	3	5	7
		8	10	12	14	9	11	13	15

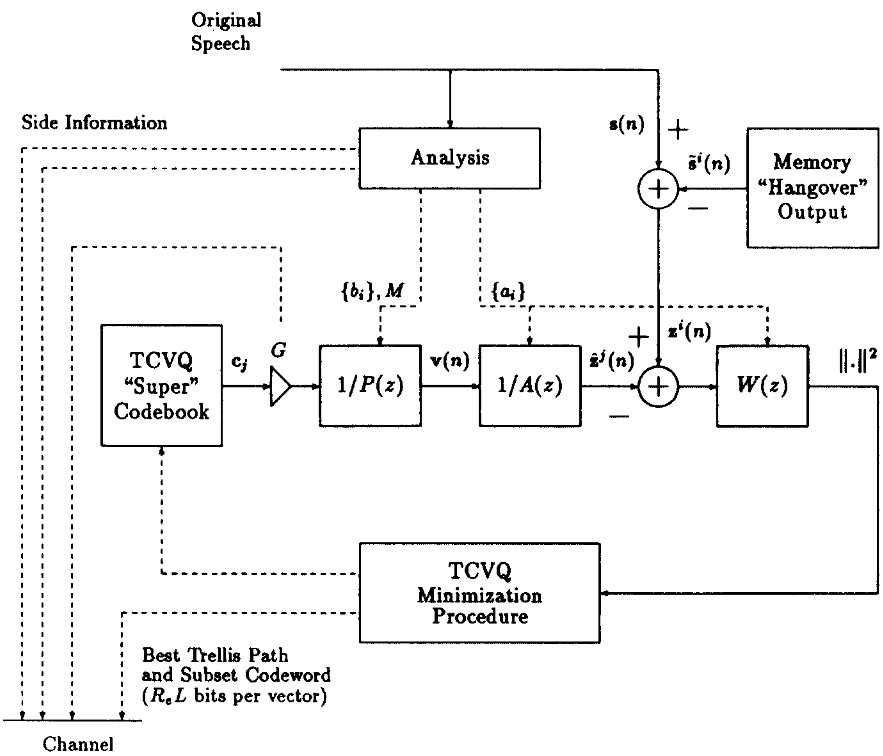


Fig. 1. Block diagram of the TCELP encoder.

Ⅲ. TRELLIS CODE-EXCITED LINEAR PREDICTION

The TCELP coder is a hybrid speech coder which uses the analysis-by-synthesis approach with trellis excitation. The basic structure of the TCELP coder is shown in Fig.1. The short-term predictor $A(z)$ is described as

$$A(z)=\sum_{k=1}^{p} a_k z^{-k}, \tag{1}$$

where $a_1, a_2, \ldots, a_p$ are the pth order LPC parameters. The long-term predictor $P(z)$ is represented by

$$P(z)=b_1 z^{-(M_p-1)}+b_2 z^{-M_p}+b_3 z^{-(M_p+1)}, \tag{2}$$

where b_1, b_2, and b_3 are three predictor coefficients and M_p describes a pitch delay in the range 16-143 samples. The speech production model includes two adaptive cascaded LPC synthesis filters $1/A(z)$ and $1/P(z)$, a TCVQ "super" codebook of excitation vectors c_j, and a gain term G. A transfer function appropriate for the weighting filter is $W(z)=A(z)/A(z/\gamma)$ [5], where $A(z)$ includes the quantized predictor parameters and $0 \le \gamma \le 1$. The purpose of the perceptual weighting filter $W(z)$ is to shape the spectra of the noise signal so that it is similar to spectra of the speech signal, thus the masking effect of the human ear [6]. Once the prediction coeefficients and pitch period are found for each frame and encoded, the original speech vectors $s(n)$ within that frame are encoded. Suppose an N-state trellis has been searched using the Viterbi algorithm to a time index of $n-1$. The memory hangover vector of the cascaded filters $1/P(z)$ and $1/A(z)$

$$\tilde{s}^i(n)=[\tilde{s}^i_1(n), \tilde{s}^i_2(n), \ldots, \tilde{s}^i_L(n)]^T$$

is computed based on the survivor path ending at node i at time $(n-1)$. There are N different vectors, one for each trellis state. Assuming the pitch delay M_p is such that $M_p \le L+1$, the memory hangover vector of the pitch predictor is described as

$$\tilde{v}^i(n)=[\tilde{v}^i_1(n), \tilde{v}^i_2(n), \ldots, \tilde{v}^i_L(n)]^T \tag{3}$$

with

$$\tilde{v}^i(n)=\sum_{j=-J}^{J} b_j v^*_j, \tag{4}$$

where

$$v^*_j=\begin{cases} \hat{v}^i_{l-l^*}(n-k^*) & \text{for } l > l^* \\ \hat{v}^i_{L-l^*+l}(n-k^*-1) & \text{for } l \le l^*, \end{cases}$$

$k^*=\lceil \frac{M_p+j}{L} \rfloor$, $l^*=M_p+j-k^*L$, and the vector sequence $\{\hat{v}^i(n-k)\}$ is the encoded version of the pitch predictor output, related to the survivor path ending at node i at time $n-1$. The memory hangover vector $\tilde{s}^i(n)$ of the cascaded filters $1/P(z)$ and $1/A(z)$ is then written as

$$\tilde{s}^i(n)=\begin{cases} \sum_{j=1}^{l-1} a_j \tilde{s}^i_{l-j}(n)+\sum_{j=0}^{L-1} a_{j+l} \hat{s}^i_{L-j}(n-1) \\ \quad +\tilde{v}^i_l(n) \qquad \text{for } L \ge p \\ \sum_{j=1}^{l-1} a_j \tilde{s}^i_{l-j}(n)+\sum_{j=0}^{L-1} a_{j+l} \hat{s}^i_{L-j}(n-1) \\ \quad +\cdots+\sum_{j=0}^{L-1} a_{j+l+mL} \hat{s}^i_{L-j}(n-1-m) \\ \quad +\tilde{v}^i_l(n) \qquad \text{for } L < p, \end{cases} \tag{5}$$

where $m=\lceil \frac{p-l}{L} \rfloor$, and the vector sequence $\{\hat{s}^i(n-k)\}$ is the encoded version of the input speech vector sequence $\{s(n-k)\}$, related to the survivor path ending at node i at time $(n-1)$. The vector $z^i(n)$ is determined by subtracting the memory hangover output $\tilde{s}^i(n)$ from $s(n)$. $\hat{z}^j(n)$ is the reconstructed vector generated by each TCVQ subset codeword c_j scaled by the gain G. The weighted prediction error is then given by

$$\begin{aligned} e^i_j(n) &= W(z^i(n)-\hat{z}^j(n)) \\ &= z^i_w(n)-\hat{z}^i_w(n), \end{aligned} \tag{6}$$

where W is a L by L lower triangular matrix described in terms of the impulse response $w(n)$ of the weighting filter, and $z^i_w(n)$ and $\hat{z}^i_w(n)$ are the frequency weighted versions of the vectors $z^i(n)$ and $\hat{z}^j(n)$, respectively. The weighted synthesis filter output is

$$\hat{z}^i_w(n) = GHc_j, \tag{7}$$

where H is a L by L lower triangular matrix with terms determined by the impulse response, say $h(n)$, of the combined pitch, formant, and weighting filters. For a given transition branch in the optimal TCVQ subset codeword is determined to minimize the squared Euclidean distance $\|e^i_j(n)\|^2_2$. Since the vectors $z^i_w(n)$ only need be computed once for each trellis transition, their computation is only a small part of the encoder's computational complexity. The gain parameter G is obtained by computing the root mean square value of the forward prediction error for each group of L_g consecutive speech samples.

Assume an N-state trellis is used to encode to a vector index of $n-1$. Given the survivor paths ending at time $n-1$, the N survivor paths at time n can be determined as follows. Let $d^i_{n-1}(z_w, \hat{z}_w)$ be the overall distortion related to the survivor path ending at node i at time $n-1$. Assume there are 2^M branches labeled with subsets entering and leaving each node. We denote the subset associated with the branch leaving node i and entering node k as S^k_i. Let the 2^M nodes at time $n-1$ with branches entering entering node k be $i_1, i_2, \ldots, i_{2M}$. The updated survivor path ending at node k at time n is determined by finding for each branch entering state k at time n, the best subset codeword that minimizes the distortion between the weighted input vector $z^i_w(n)$ and the weighted synthesis vector $\hat{z}^j_w(n)$. Let this best codeword from subset S^k_i be c^k_i. Thus, c^k_i is the codeword c from S^k_i that minimizes

$$\|z^i_w(n) - \hat{G}Hc\|^2_2, \quad i = i_1, i_2, \ldots, i_{2M}.$$

Then, we compute the overall distortion associated with each of the 2^M possible paths to node k at time n and select the path with the minimum distortion as the updated survivor path ending at state k. Thus, the TCVQ minimization procedure computes

$$d^k_n(z_w, \hat{z}_w) = \min_{i \in \{i_1, i_2, \ldots, i_2M\}} (d^i_{n-1} + \|z^i_w(n) - \hat{G}Hc^k_i\|^2_2). \tag{8}$$

After all N survivor paths at time n are determined, the time index is incremented and the process is repeated until a certain depth corresponding to positive integer multiples of the vector dimension, L. Then, M bits per vector to specify the best trellis path and R_eL-M bits per vector to indicate the subset codeword are transmitted. In the receiver, the transmitted data produce a sequence of codewords. Each codeword is scaled by $\hat{G}$, the corresponding quantized gain. The resulting signal is passed through the pitch and formant predictors to produce the reconstructed version of the input speech vector $s(n)$.

Ⅳ. TCELP OPTIMAL CODEBOOK DESIGN

In this section, we will introduce a procedure for designing the optimal "super" codebook, subject to the TCELP structure, by applying the generalized Lloyd algortihm [7] to a training sequence of input vectors $z(n)$. For each optimization itration, the updated codebook is optimal for the current input sequence in the sense that the perceptual weighted distortion between the input vectors $z(n)$ and the reconstructed vectors $\hat{z}(n)$ is minimized. However, since any change of the codebook alters the input sequence, convergence of the design can not be assured.

As an initial "super" codebook, we use a random codebook in which each codeword is constructed of samples of unit-variance white Gaussian process. We chose the Gaussian distribution since the probability density function of the prediction error sequece (after both formant and pitch predictions) is reasonably modeled as white and Gaussian [8]. Given the initial codebook, the initial subsets are formed based on Ungerboeck's set partitioning method [9].

The codebook optimization algorithm of the TCELP encoder is described as follows and is similar to the algorithm for structure 1 optimization in [3]. Given a training input sequence, χ and an initial codebook, S^1, set $k=1$, and at iteration k:

1. Encode the training sequence using the Viterbi algorithm and the TCVQ structure with codebook S^k. Denote the resulting distortion as

$$E(k)=\frac{1}{\|\chi\|}\sum_{S(n)\in\chi}\|z_w(n)-\hat{G}Hc_j^k\|_2^2 .$$

2. Partition the input vectors $z(n)$ associated with the training sequence into sets Q_j^k, so that $z(n)\in Q_j^k$, if and only if its weighted version $z_w(n)$ was encoded as $\hat{G}Hc_j^k$.

3. Update the TCVQ codebook as S^{k+1} by

$$c_j^k=\Big(\sum_{z(n)\in Q_j^k}\hat{G}^2H^TH\Big)^{-1}\sum_{z(n)\in Q_j^k}\hat{G}H^Tz(n).$$

Set $k=k+1$ and go to step 1.

Since the input vectors $z(n)$ computed from the speech signal are dependent on the current excitation codebook, $E(k)$ is not guaranteed to decrease monotonically with k. Typically, a large decrease of E is obtained in the first few iterations. The optimization process can be halted by a suitable termination criteria. In this paper, the process is stopped after 12 iterations. The codebook S^k generating the minimum distortion is selected as the "optimal" TCVQ codeook.

V. SYMBOL RELEASE RULE

If we search the entire trellis before releasing any symbols, the best performance can be achieved. However, this search introduces excessive encoding delay, and a trellis symbol release rule corresponding to a suboptimum strategy is required in a practical speech coding system.

The symbol release rule considered herein is similar to that in [3] and is described as follows. Let $K_r\geq 1$ and $K_d\geq 0$ define, respectively, the number of branch symbols released and the depth of the trellis search at which a hard decision is made. Suppose the trellis encoding has proceeded to a sample $n=jK_r$, j an integer. The survivor path with the minimum distortion is traced back $K_r=K_d$ branches, and the K_r branch symbols (codewords) corresponding to samples $jK_r-K_d-K_r$, ..., jK_r-K_d-1 are released. Define the node that the best survivor path at sample jK_r passes through at sample jK_r-K_d as $z^*(jK_r-K_d)$. Each survivor path at sample jK_r is traced back to sample jK_r-K_d. If the resulting node is not $z^*(jK_r-K_d)$, then the associated survivor metric at sample jK_r is set to ∞. If the resulting node is $z^*(jK_r-K_d)$, then no change is made. This effectively "prunes" all survivor paths that would lead to an inconsistent trellis path. The performance of the coder is expected to increase at the expense of longer encoding delay as K_r or/and K_d increase. Hence, K_r and K_d are generally selected as large as permissible in each application.

VI. Encoding the Side Information

Referring to Fig. 1, the TCELP coding system encodes three different parameters (gain, formant, and pitch parameters) as the side information. The quantization levels for the scalar quantization of gain parameters were designed by applying the generalized Lloyd algorithm to a training sequence. A low distortion representation of the formant coefficients in important for a good coder performance. For high quality encoding, the formant coefficients are first transformed to LSP parameters [10], and then the backward sequential adaptive quantization scheme (AQBW) [11] is applied to quantize the LSP parameters. The pitch coefficients are treated as vectors and encoded by vector quantization techniques. The VQ codebook is determined by applying the generalized Lloyd algorithm to a training sequence. The designed codeword is stabilized (if necessary) using a method in [12]. The pitch period is noiselessly encoded using 7 bits/period.

Ⅶ. COMPLEXITY

In this section, the encoding complexity of the TCELP system is evaluated. The analysis concentrates only on the trellis excitation part of the system. The computational requirements related to determining and encoding both formant and pitch predictors are excluded since those are negligible in comparison to the complexity required in the trellis excitation portion, and each of these operations is common to most CELP systems using the adaptive forward prediction. Through this section, we assume enough memory to store all intermediate quantities used in determining the optimal codeword.

The TCELP encoder operates on L-dimensional source vectors, with a p *th* order formant predictor, a q *th* order pitch predictor, and an N-state trellis with 2^M branches entering and leaving each trellis state. Computing $\mathbf{z}^i(n)$ requires roughly $NL(p+q)$ additions and $NL(p+q)$ multiplies per source vector. Forming $\mathbf{z}_w^i(n)(=\mathbf{W}\mathbf{z}^i(n))$ requires another $NL(L-1)/2$ additions and $NL(L-1)/2$ multiplies per source vector. The number of computations required to compute $\mathbf{H}\mathbf{c}_j$, $j=1, 2, \ldots, \|S\|$, once per frame is roughly $L(L-1)\|S\|/2l_f$ additions and $L(L-1)\|S\|/2l_f$ multiplies per source vector, where l_f is the number of vectors in a synthesis frame. Computing $\hat{\mathbf{z}}_w^j(n)=\hat{G}\,\mathbf{H}\mathbf{c}_j$ requires another $L\|S\|/l_g$ multiplies per source vector, where l_g is the number of vectors in the gain update rate.

In TCVQ, there are $2^{\hat{R}L+M}$ subsets, each consisting of exactly 2^{R_eL-M} codewords. Given the vectors $z_w^i(n)$ and $\hat{z}_w^j(n)$, a full search to find the optimal codeword in each subset that minimizes $\|e_j^i(n)\|_2^2$ requires $L2^{R_eL-M}$ additions, $L2^{R_eL-M}$ multiples, and $2^{R_eL-M}-1$ two-way comparisons per subset. Thers are 2^M entering branches in each trellis state, with branch metrics the corresponding subset MSE. Thus, computing the $N2^M$ branch distortions requires, $NL2^{R_eL}$ additons, $NL2^{R_eL}$ multiplies, and $N(2^{R_eL}-2^M)$ two-way comparisons per source vector. Determining the best survivor path at each trellis state requires an additional $N2^M$ additions and $N(2^M-1)$ two-way comparisons per source vector. The total complexity considered so far is then approximatoly

$$
\begin{aligned}
&NL(p+q+2^{R_eL})+\frac{NL(L-1)}{2}\\
&+\frac{L(L-1)\|S\|}{2l_f}+N2^M \text{ additions,}\\
&NL(p+q+2^{R_eL})+\frac{L(L-1)\|S\|}{2l_f}\\
&+\frac{NL(L-1)}{2}+\frac{L\|S\|}{l_g} \text{ multiplies,}\\
&N(2^{R_eL}-1) \text{ two-way comparisons.}
\end{aligned}
\tag{9}
$$

The computational burden related to the symbol release rule was not included in (9).

The full search excitation encoding complexity of L'-dimensional vector excitaion coding is approximately

$$
\begin{aligned}
&L'2^{RL'} \text{ additions,}\\
&L'2^{RL'} \text{ multiplies,}\\
&2^{RL'}-1 \text{ two-way comparisons.}
\end{aligned}
\tag{10}
$$

For the same dimension of excitation vectors, TCVQ is more complex, but has better performance [3]. Hence, it is possible that for equal distortion systems, the TCELP encoder may be less complex than the vector excitation coding. To reduce the computational burden related to the subset search, a suboptimal method using tree-searched VQ is possible [3].

In the open-loop gain computation, the complexity to computing and encoding the gains, G, is negligible in comparison to the computation

Table 2. Korean sentences Used to Evaluate Encoding Performance.

1. 미는 피부 한 겹질의 차이 입니다.	(Female)
2. 지나친 흡연은 건강을 해칩니다.	(Female)
3. 이번겨울은 예년과 달리 포근합니다.	(Male)
4. 과학기술은 경제발전의 원동력입니다.	(Male)
5. 일에서 십까지의 합은 오십오입니다.	(Male)
6. 어제 산 물건이 벌써 고장이 났다.	(Male)
7. 올림픽은 전 인류의 축제입니다.	(Female)

Table 3. English sentences Used to Evaluate Encoding Performance.

1. The pipe began to rust while new.	(Female)
2. Add the sum to the product of these three.	(Female)
3. Oak is strong and also gives shade.	(Male)
4. Thieves who rob friends deserve jail.	(Male)
5. Cats and dogs each hate the other.	(Male)
6. Almost everything involved making the child mind.	(Male)
7. The trouble with swimming is that you can drown.	(Female)

Table 4. Parameter values for the 8 kbps TCELP Coder.

parameters	values
encoding rate of excit. seq., R_e	58 (bits/sample)
rate expansion factor, $\tilde{R}$	14 (bits/sample)
dimension of excit. vector, L	8 (samples)
number of trellis states, N	8
M	1
K_r, K_d	40, 40 (samples)
excitation gain	4 bits/5 ms
formant predictor ($p=10$)	31 bits/20 ms
pitch predictor (3 tap)	6 bits/20 ms
pitch period (16-143)	7 bits/20 ms
weighting factor, γ	0.8
number of design data samples	76, 480
number of optimization iterations	12

requirements given in (9), and is excluded this section.

Ⅷ. EXPERIMENTAL RESULTS

In this section, we evaluate the effectiveness of TCELP coders at low bit rates. Both Korean and English sentences of Tables 2 and 3 are used to evaluate encoding performance. In each language, Sentences 1, 3, 6, and 7 were used to design the coder, and sentences 1-5 were used to evaluate the coder performance. Table 4 lists the parameter values for the TCELP coder operating at 8 kbps.

Table 5 presents the simulation results for the unweighted and weighted 8 kbps TCELP coders. The simulations using Korean sentences produced the average SNR and SEGSNR of 15.68 and 14.97 dB, respectively. The simulations using English sentences produced the average SNR and SEGSNR of 15.56 and 14.35 dB, respectively. To assess the subjective quality of the TCELP encoder, a TCELP reconstructed sentence was compared with that of a μ-law PCM system [13]($\mu=255$) operating at bit rates of 3 through 8 bits per sample. Informal listening tests indicate the 8 kbps TCELP system performs roughly between the 6-bit and 7-bit μ-law PCM with $\mu=255$. Also it revealed that the advantage of error weighting is small, but can be heard. For example a "warble" noise in the word "rob" of the English sentence 4 was reduced with the weighting filter. An empirically "optimal" value of γ was found to be 0.8.

Table 5. The Performance of the 8 kbps TCELP Coder.

		SNR/SEGSNR (dB)				
		Sentence Number				
	γ	1	2	3	4	5
Korean	1	16.63/16.22	16.31/16.23	15.77/14.12	14.91/13.87	14.78/14.39
	0.8	15.52/15.32	15.33/15.29	14.78/13.04	14.30/13.38	14.10/13.74
English	1	17.36/15.61	17.24/15.38	14.02/12.22	13.93/13.94	15.25/14.62
	0.8	16.66/14.96	16.51/14.88	13.63/11.70	13.47/13.44	14.31/14.02

Ⅸ. CONCLUSIONS

An effective 8 kbps vocoder for digital cellular system, called trellis code-excited linear prediction coding (TCELP), was introduced, which incorporates TCVQ [3] in the CELP structure [1]. We formulated such a combination of TCVQ and CELP. A training sequence-based algorithm was then introduced for iteratively designing the optimal codebooks subject to the TCELP structure. Also, we described the encoding complexity and trellis symbol release rules. Then, simulation results for the efficient 8 kbps TCELP coder was presented in terms of SNR/SEGSNR and the informal listening tests.

참 고 문 헌

1. M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP) : Highquality speech at very low bit rates," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1985, pp.252-255.
2. M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol.38, pp. 82-93, Jan. 1990.
3. T. R. Fischer, M. W. Marcellin, and Min Wang, "Trellis coded vector quantization," *IEEE Trans. Inform. Theory*, vol.IT-37, pp.1551-1566, Nov. 1991.
4. G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE(Invited Paper)*, vol.61, pp.268-278, Mar. 1973.
5. B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-27, pp.247-254, June 1979.
6. M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Amer.*, pp.1647-1652, Dec. 1979.
7. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol.COM-28, pp.84-95, Jan. 1980.
8. B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol.COM-30, pp.600-614, April 1982.
9. G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol.IT-28, pp.55-67, Jan. 1982.
10. F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol.57, supplement no.1, S35(A), 1975.
11. N. Sugamura and N. Farvardin, "Quantizer design in LSP speech analysis-synthesis," *IEEE Journ. Selected Areas in Commun.*, vol.6, pp.432-440, Feb. 1988.
12. R. P. Ramachandran and P. Kabal, "Pitch prediction filters in speech coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-37, pp.467-478, April 1989.
13. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Englewood Cliffs, NJ : Prentice-hall, Inc., 1984.

姜 尙 遠(Sangwon Kang) 정회원

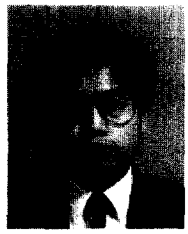
1980년 2월 : 한양대학교 전자공학과 졸업(공학사)

1982년 2월 : 서울대학교 전자공학과 대학원 졸업(공학석사)

1990년 8월 : 미국 Texas A&M Univ. 전기공학과 졸업(공학박사)

1982년 12월 ~ 현재 : 한국전자통신연구소 이동통신연구단 신호처리연구실장

※주관심분야 : 이동통신, 음성신호처리, 통신이론



李 寅 誠(Insung Lee) 정회원

1960년 6월 8일생

1983년 2월 : 연세대학교 전자공학과 졸업(공학사)

1985년 2월 : 연세대학교 전자공학과 대학원 졸업(공학석사)

1992년 12월 : 미국 Texas A&M Univ. 전기공학과 졸업(Ph.D.)

1986년 5월 ~ 1987년 7월 : 한국통신 연구개발단 전임연구원

1993년 2월 ~ 현재 : 한국전자통신연구소 신호처리연구실 선임연구원

※주관심분야 : 음성, 영상데이타 압축, 적응필터, 이동통신

韓 敬 浩(Kyongho Han) 정회원

1959년 6월 25일생

1982년 2월 : 서울대학교 전자공학과 졸업(공학사)

1984년 2월 : 서울대학교 전자공학과 대학원 졸업(공학석사)

1992년 8월 : 미국 Texas A&M Univ. 전기공학과 졸업(Ph.D.)

1985년 5월 ~ 1987년 12월 : 한국전기통신공사 품질보증단 전임연구원

1992년 10월 ~ 1993년 8월 : 한국전자통신연구소 신호처리연구실 선임연구원

1993년 8월 ~ 현재 : 단국대학교(서울) 전기공학과

※주관심분야 : 컴퓨터구조, VLSI, 마이크로프로세서응용.