Improving EVRC Half Rate by the Algebraic VQ-CELP

Fenghua Liu and Ryan Heidari

Nokia Mobile Phones
9605 Scranton Road #150
San Diego, CA 92121, USA

ABSTRACT

This paper presents an algebraic vector quantized codebook excited linear prediction (AVQ-CELP) speech codec. The objective is to enhance the half rate mode of IS-127, the enhanced variable rate codec (EVRC). In AVQ-CELP scheme, only the perceptually important components are encoded, and the selection of the components is done in a way similar to the ACELP. An open-loop procedure is used to select the sub-vectors. The selected sub-vectors are concatenated and vector quantized. An analysis-by-synthesis strategy is used to determine the optimal excitation. The generalized Lloyd algorithm (GLA) is used to optimize the AVQ codebook. In order to improve the synthesis quality of voiced frames, a two-pulse version of ACELP is used in the strong voiced frames. The proposed algorithm was incorporated in the Nokia CDMA handset prototype. Under a joint collaboration effort with SK Telecom, a field-testing was performed in Korea to evaluate the performance of the proposed AVQ algorithm. The results indicate a considerable improvement relative to the standard EVRC operating at the maximum half-rate.

1. INTRODUCTION

The enhanced variable rate codec (EVRC) is a new wireless standard (IS-127)[2] for CDMA technology. EVRC is based on ther elaxation CELP (RCELP) paradigm[3,4]. Unlike conventional CELP codec, RCELP attempts to match a modified speech residual signal generated by a time-warped version of the original residual that conforms to a simplified pitch contour. As a result, the pitch information is transmitted over a frame instead of a subframe. Consequently, more bits are allocated to the fixed codebook encoding and the error detection/protection. The EVRC speech quality is better than G.728 and comparable to the QCELP-13 (IS-733), which operates at peak rate of 13.3kbit/s.

A low bit-rate speech codec is highly desirable in wireless applications because more users can be accommodated. The new challenge in the speech coding is to develop a toll-quality speech codec at 4 kbits/s or lower. The ITU-T has initiated to investigate the possibility of encoding speech at 4 kbits/s for universal application.

In general, low bit rate speech coders extract a set of parameters to describe the process of the speech generation, and transmit these parameters instead of the speech waveform. Typically, there are two types of parameters that are encoded and transmitted: The model parameters, i.e., the linear prediction coefficients (LPC) and the excitation parameters. The LPC can be transformed into the line spectrum pairs (LSP) and 20-24 bits are sufficient to encode LSP. The remaining problem is how to encode the excitation signal

Linear-prediction based on analysis-by-synthesis (LPAS) methods have led to efficient speech coding at rates between 4 and 16kbit/s. In the LPAS process, the optimal set of parameters for reproducing each segment of the original speech signal is found at the encoder. Then, the quantized parameters are transmitted to the decoder that uses an identical speech production model and the set of parameters to synthesize the speech waveform.

After we review the conventional CELP excitation in section 2, we introduce the algebraic VQ CELP in section 3. We show a strategy to determine the position of the sub-vectors to be quantized, fast search for the AVQ codebook and a target shifting procedure to increase the efficiency of the VQ. Multi-mode excitation is introduced to improve the synthesized speech quality of the voiced speech frames. The AVQ codebook design is presented in section 4. The quality of the AVQ-CELP codec is reported in section 5, where tests were performed in an actual field-testing environment.

2. ALGEBRAIC VQ CELP

The well-known methods to represent the excitation include the multi-pulse linear prediction coding (MPLPC), code-excited linear prediction (CELP), and algebraic CELP (ACELP).

In code excitation linear prediction (CELP) system, the excitation vector is chosen from a set of pre-stored stochastic sequences. During the codebook search, all possible codevectors from the codebook are passed through a pitch filter and a synthesized filter. The codevector that generates the minimum mean-squared error is chosen as the desired excitation. Since identical codebooks are stored at both encoder and decoder, only the index corresponding to the selected codevector is transmitted.

In the multi-pulse excitation linear prediction coder (MPLPC)[1], no voiced/unvoiced classification is performed. The excitation is specified by a small set of pluses, each with different amplitudes and locations. Since there is no constraint on the pulse positions and amplitudes, this algorithm requires a large number of bits to encode the pulse positions and amplitudes. Therefore, structured pulse positions and amplitudes would be more desirable in terms of the bit saving reduction of the complexity.

Algebraic code excited linear prediction (ACELP)[5] scheme uses an interleaved single-pulse permutation design to divide the pulse positions into several tracks. All pulses have fixed amplitudes and only the sign of the pulses are transmitted. Incorporating the fast deep-tree search and pitch shaping, ACELP is very successful in providing the high quality speech at low bit rate. Recent speech coding standards used in wireless technologies for TDMA, CDMA and GSM are all based on the ACELP.

ACELP uses an efficient way to encode the pulse positions and signs. In order to maintain a good quality, only four to eight pulses are used depending on the subframe length. In a low-bit rate codec, there are not enough bits to encode the excitation pulses. Therefore, the number of excitation pulses must be reduced or the pulse positions are constrained to some pre-selected positions, which generally results in a quality degradation of the synthesized speech. Table 1 shows the bit allocation for the EVRC full-rate and half-rate modes.

Table 1. The bit allocation of the EVRC full-rate and half-rate

Parameter	Full-Rate	Half-Rate
Spectral Indicator	1	-
LSP	28	22
Pitch Delay	7	7
Delta Delay	5	-
ACB Gain	3x3	3x3
FCB Shape	3x35	3x10
FCB Gain	3x5	3x4

The most considerable bit reduction from the full-rate to the half-rate is in the fixed codebook shape, which reduces from 35 bits to 10 bits. In the full-rate EVRC, 35 bits are used to encode the 8 excitation pulses, whereas in half-rate, only 10 bits are used to encode 3 excitation pulses. The insufficient number of excitation pulses is the main cause of quality degradation in the half-rate EVRC.

The main task in improving the performance of the low-bit rate codec is to increase the coding efficiency. A good choice is to use vector quantization (VQ) to represent the excitation. In our approach, we generalize the multi-pulse excitation concepts to the multi-sub-vector. Several samples are grouped into a sub-vector; therefore, there are several sub-vectors in each subframe. Only the perceptually important sub-vectors are encoded, other unselected sub-vector elements are set to zero. The position of the sub-vectors is encoded in a way similar to the algebraic codebook. The new speech coding method is called the algebraic VQ-CELP. Fig.1 shows the AVQ-CELP scheme.

In the algebraic VQ-CELP, the residual signal generated by passing the target vector through the linear predictor (LP) is first filtered by a pitch-filter to remove the long-term correlation

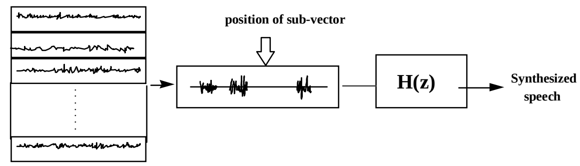

Fig.1 The Algebraic VQ CELP

existing in each sub-frame. Five samples are grouped together to form a sub-vector. In order to maintain the pitch periodicity of the fixed excitation, the partition of the sub-vectors is based on the pitch period. The total number of the sub-vectors is equal to the integer part of pitch/5 and is bounded by 3 and 11. The sub-vectors are arranged in an interleaved order as shown in table 2.

Table 2. The allocated sub-vector locations

No. of Bit	Sub-vector locations			
0	0	3	6	9
1	1	4	7	10
2	2	5	8	(11)

Six bits are used to present the positions of the three sub-vectors. This technique is a direct extension of the multi-pulse, where only one sample is selected to be quantized. Here, we select a vector to quantize instead of a scalar.

Typically, speech is classified as voiced/unvoiced. For different speech waveforms, we can use different modes for encoding. Since the adaptive codebook can not remove all the redundancy, the pitch period excitation can provide the improvement on the synthesized speech. The selection of the three sub-vectors is based on the pitch period for the voiced speech. In the unvoiced case, the selection of three sub-vectors is always based on the subframe length. In a strong voiced case, a two-pulse ACELP scheme is used instead of the VQ. Switching between modes is done based on the gain of adaptive codebook; therefore, no extra bit is needed to indicate the mode selection.

In AVQ-CELP, the first step is to select three perceptually important sub-vectors. There are two ways to select these sub-vectors. One is the closed-loop approach, where each possible combination of the three sub-vectors is passed through the synthesized filter. The combination of the three sub-vectors which results in minimum mean squared-error is selected. By this way, the selection of the sub-vectors and the codevectors are jointly optimized. The joint operation may result in high complexity. In order to reduce the complexity, we can use an open-loop approach instead, where the selection of the sub-vectors and the codebook search are sequentially performed. In this approach, the selection of the sub-vectors is based on the original residual signal. Full search is used to select the three sub-vectors. In each selection process, the three sub-vectors chosen according to the interleaved order, are kept the same and the other unselected sub-vectors elements are set to zero. The resultant vector is passed through a pitch shaping and a synthesized filter to generate the synthesized signal, which is compared to the target vector. The three sub-vectors that results in the minimum distortion are selected and quantized. Fig.2 shows the basic procedures to determine the three sub-vectors.

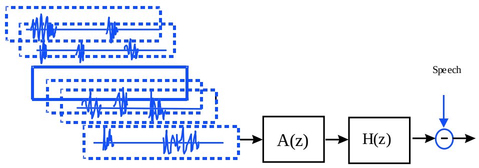

Fig.2 Determination of the Three Sub-Vectors

Selection of three perceptually important sub-vectors enables us to quantize the excitation vector more efficient and use less memory to store the fixed codebook. The three selected sub-vectors are concatenated to form a new 15-dimensional vector, which is quantized based on the closed-loop analysis.

In the conventional CELP synthesis process, the codebook is searched directly from the target vector. In EVRC, when the speech codec rate changes from full-rate to half-rate, the number of bits used to present the excitation is reduced from 35 to 10. Therefore, we do not have enough excitation patterns to match the original excitation waveform. We need an alternative approach to regenerate the excitation patterns based on the available information. One approach is to increase the patterns of the fixed codebook by circularly shifting the fixed-codebook based on the signal generated by the adaptive codebook. Ideally, the selection of the shift should be done on the target vector. But doing so, it would require additional bits to transmit the circular shift information. However, we can avoid transmitting the shift information by using the adaptive codebook as a reference signal. The circular shift operation is performed only for the voiced speech signal. The shift decision is determined based on the gain of the adaptive codebook. If the adaptive codebook gain is above a certain threshold, the adaptive codebook tracks the input speech very well, then the circular shift is performed, otherwise, no shift operation is carried out. Ideally, the selection of the shift should be done on the basis of the analysis-by-synthesis. However, this operation will increase the complexity. The open-loop strategy is used to decide the shift of the codebook to maximize the cross-correlation of the target signal and the excitation signal.

$$\max\{(|c^T H^T H x_a|)\}$$

where x_a is the adaptive codebook contribution. Since the decision of the shift is based on the adaptive codebook, the shift information is not transmitted. The receiver just needs to repeat the same search process for the given shift index.

A pitch-filter is used to improve the perceptual quality of the synthesized speech. It turns out that pitch shaping of the excitation can be incorporated in the codebook search, which is equivalent to modify the impulse response of weighted synthesized filter.

In a typical CELP coder, the adaptive and fixed codebooks are sequentially searched in order to reduce the complexity. The adaptive codebook is usually searched first. Then the contribution of the adaptive codebook is subtracted from the input speech waveform. The resulting signal is called the target signal. The contribution of our approach is related to forming a new fixed codebook based on vector quantization. The detailed fixed codebook search is shown in Fig. 3.

The input of the fixed-codebook search is the target vector $x(n)$, its corresponding residual signal $x_w(n)$, which is obtained by filtering the target vector through inverse filter $1/H(z)$, the adaptive codebook $x_a(n)$, and the $h(n)$, thr impulse response of the weighted synthesized filter, the adaptive codebook gain $\boldsymbol{g_a}$ and the pitch P, which are determined during adaptive codebook search. The signal $x_a(n)$ is first passed through a long-term de-correlation filter $1\text{-}g_a z^{-P}$.

The adaptive codebook can not remove all the long-term correlation in the low bit rate CELP coder. As a result, the fixed codebook contribution should contain some periodicity. The impulse response h(n) is passed through a pitch-shaping filter to enhance the periodic property of the synthesized filter. From the pitch-shaping impulse response h(n), we can construct an impulse response matrix H , where H is a Toeplitz matrix

$$H=\begin{bmatrix} h[0] & 0 & . & . & . & . & 0 \\ h[1] & h[0] & 0 & & & & 0 \\ h[2] & h[1] & h[0] & 0 & & & . \\ h[3] & h[2] & h[1] & h[0] & 0 & & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ h[n-1] & h[n-2] & h[n-3] & . & . & h[0] & \end{bmatrix}$$

From the impulse response matrix, the auto-correlation matrix $\Phi = H^T H$ is computed, the auto-correlation matrix is used to determine the best codevector.

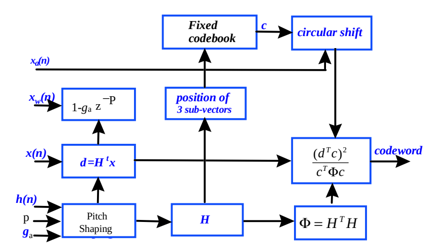

Fig. 3 Codebook Search in AVQ-CELP

A backward-filtered signal d, $d= H^t x$, is first computed, which determines the positions of the three sub-vectors and the codebook search. The criterion to select the three sub-vectors is based on maximizing the correlation between the backward-filtered signal, d, and the residual signal $x_w(n)$. Based on the positions of the three sub-vectors, an excitation vector can be constructed for every codevector. The resulting excitation vector is circularly shifted and the correlation between the circular-shifted excitation vector and the $x_a(n)$ is calculated.

The shift, which generates the maximum correlation, is selected as the final shift value. The excitation codevector, along with d and Φ is used to select the optimal codevector which maximizes

$$\frac{(c^T d)^2}{c^T \Phi c}$$

3. CODEBOOK DESIGN OF AVQ CELP

The Generalized Lloyd algorithm (GLA) is used to design the VQ codebook. The mean squared error (MSE) associated with the *j-th* codevector c_j can be expressed as

$$\varepsilon = \sum_i \|t_i - g_i H_i P_i S_i c_j\|^2$$

where t_i , the target vector for the fixed codebook search, is obtained by subtracting the zero input response and the contribution of the adaptive codebook. *Hi*, which denotes the impulse response matrix, *Pi,* which denotes the mapping from the selected subvector to the excitation location, and *Si* denotes the shift operation on the fixed codevector *cj*. The reference signal for shifting is the adaptive codebook. Following example shows a shift by one matrix.

$$s = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & .. & 0 & 0 \\ .. & . & . & . & \ldots & . & . \\ . & . & . & . & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The optimal codevector that minimizes the MSE is given by

$$c_j = \left(\sum_i \left(g_i^2 S_i^T P_i^T H_i^T H_i P_i S_i \right) \right)^{-1} \sum_i \left(g_i S_i^T P_i^T H_i^T t_i \right)$$

The following two steps are used to design the AVQ codebook.

1. Cluster the target signal into N set, each cluster is associated with the codevector c_j.

2. Find the optimal codebook c_j based on the optimal partitions.

After some iterations, the codebook will converge. It is very interesting to note that some trained codevectors have pulse like shape, which is associated with the voiced speech and some have noise like shape, which is associated with the unvoiced speech.

4. PERFORMANCE EVALUATION

After several informal MOS tests, the proposed AVQ-CELP algorithm was first converted to the fixed-point C simulation. The code was further implemented on a real-time DSP platform. The AVQ-CELP half-rate code was integrated with the standard EVRC code, since only the fixed-codebook section of EVRC was modified. A private service option is selected because the modified EVRC is not interoperable with the standard EVRC codec. The DSP code is integrated to the NOKIA 2180 CDMA handset.

The field-testing was performed in Seoul, Korea as a joint collaboration with SK Telecom. In the field testing, the proposed EVRC half-rate code was tested against the standard EVRC handset running at the maximum half-rate under background noise (street noise) and frame error conditions. The preliminary test results indicate that our modified EVRC half-rate codec achieved considerable improvement over the standard EVRC half-rate codec.

5. ACKNOWLEDGEMENT

We would like to express our sincere gratitude and thanks to Dr. Joo Sik Lee and Mr. Woosung Kwon of SK Telecom for their co-operation and assistance in our field testing in Korea and helping us evaluating the performance of the Nokia modified EVRC half-rate codec.

REFERENCES:

[1] Atal, B. S. and Schroeder, M. R., “Stochastic Coding of Speech at Very Low Bit Rates”, Proc Int. Conf. Comm., pp 1610-1613, Amsterdam, 1984.

[2] Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems, IS-127, September 9, 1996

[3] High Rate Speech Service Option 17 for Wideband Spread Spectrum Communication Systems, IS 733, 1998

[4] Kleijn, W. B., Kroon, P., and Nahumi, D., “The RCELP Speech-Coding Algorithm”, European Transactions on Telecommunications, Vol 5, Number 5., pp 573-582, Sept/Oct 1994.

[5] Nahumi, D and Kleijn, W. B., “An Improved 8 kb/s RCELP Coder”, IEEE Workshop on Speech Coding, 1995.

[6] Salami, R., Laflamme, C., Adoul, J-P., et.al, “A Toll Quality 8 kbits Speech Coder for Personal Communication Systems (PCS)”, IEEE Trans. on Vehicle Technology, 1994.

[7] McAulay, R. J. and Quatieri, T.E., “Low-rate Speech Coding Based on the Sinusoidal Model”, in *Advances in Acoustics and Speech Processing,* Sondhi, M. and Furui, S. Eds., Marcel-Dekker, pp. 165-207, New York, 1992.

[8] Kleijn, W.B. and Haagen, J., “Transformation and Decomposition of the Speech Signal for Coding”, pp. 249-252, *IEEE Signal Processing Letter*, 1994.

[9] McCree, A., Truong, K., George, E., Barnwell, T., and Viswanathan, V., “A 2.4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard”, *Proc. ICASSP96*

[10] Oyha, T., Suda, H., and Miki, T., “5.6 kb/s PSI-CELP of the Half-Rate PDC Speech Coding Standard”, Proc. *IEEE Vehicular Technol*. Conf. June. 1994, pp. 1680-1684

[11] Jang. Y., Cuperman, V., “An improved 2.4kbps Class-Dependent CELP Speech Coder”, pp. 1414-1417, *IEEE ICC95.*