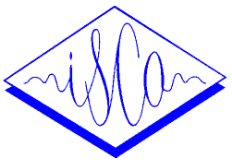

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

EFFICIENT HARMONIC-CELP BASED HYBRID CODING OF SPEECH AT LOW BIT RATES

Yong-Soo Choi, Sueng-Kyun Ryu+, Young-Cheol Park+, and Dae-Hee Youn+*

* Digital Network R&D Lab., LG Information & Communications, Ltd.
60-39 Kasan-dong, Kumchun-ku, Seoul 153-023, KOREA
+ Center for Signal Processing Research, Yonsei University
134 Shinchon-dong, Sudaemun-ku, Seoul 120-749, KOREA
E-mail: * cando@lgic.co.kr, + [syryu@radar., young@lethe., dhyoun@]yonsei.ac.kr

ABSTRACT

This paper presents an efficient Harmonic-CELP hybrid coder at 2.4 kbps utilizing the well-known characteristics of the Harmonic and CELP coders. According to frame voicing decision, the proposed hybrid coder switches the RP-VSELP coder as a fast CELP in case of unvoiced, or an improved Harmonic coder in case of voiced. The proposed Harmonic-CELP hybrid coder has several features as follows: fast CELP coding, fast harmonic estimation, variable dimension harmonic vector quantization, perceptual weighting including Bark frequency resolution, fast harmonic synthesis, and naturalness control by band voicing. To demonstrate the performance of the proposed hybrid coder, a 2.4 kbps coder has been implemented and compared with 5.3 kbps ACELP and 4.4 kbps IMBE as reference coders. From results of subjective tests, the proposed hybrid coder showed good quality at about half rates of the reference coders.

1. INTRODUCTION

The Code Excited Linear Prediction (CELP) [1] coding has been proved to be the most efficient technique to produce high quality speech at low bit rates as low as 8 kbps. However, the CELP quality degrades rapidly at rates below 4 kbps while the Harmonic coder such as Sinusoidal Transform Coding (STC) [2] and Multi Band Excitation (MBE) [2][3] has good quality. It has been widely reported that the CELP coders synthesize well noisy unvoiced signals and the Harmonic coders do periodic voiced signals at low bit rates below 4 kbps.

This paper presents an efficient Harmonic-CELP (EHC) hybrid coding algorithm at low bit rates based on the well-known characteristics of the Harmonic and CELP type coders. According to frame voicing, in case of unvoiced the EHC coder employs the Regular Pulse Vector-Sum Excited Linear Prediction (RP-VSELP) [4] which we previously presented as a fast CELP algorithm, and in case of voiced uses an improved Harmonic coding algorithm. As a result, the proposed coder can produce good quality with low complexity.

The RP-VSELP using regular pulse basis vectors significantly reduces the VSELP [5] complexity while maintaining speech quality and robustness to channel errors. In addition, the RP-VSELP can optimize its codebook through an iterative close-loop training process [6].

The improved Harmonic coding has several features as follows: fast harmonic estimation, referred to a delta adjustment (DA) method, perceptual weighting including Bark frequency resolution [7] and naturalness control by harmonic band voicing. A DA method using only an integer pitch is presented for a reliable fast harmonic estimation. Also the DA method significantly reduces the complexity while maintaining the spectral distance performance, compared with the typical harmonic analysis methods using fractional pitch. For harmonic vector quantization, a new perceptual weighting function which reflects both the frequency masking effect and Bark frequency resolution is introduced. To enhance naturalness of the reproduced voiced speech, a colored noisy signal is mixed with the synthetic excitation signal using band voicing values.

To demonstrate the performance of the proposed coder, a 2.4 kbps EHC coder has been implemented and compared with 5.3 kbps ACELP [8] and 4.4 kbps IMBE [3] as reference coders. From results of subjective quality tests, the proposed hybrid coder showed good quality at about half rates of the reference coders.

This paper is organized as follows. In Section 2 we describe background on Harmonic-CELP hybrid coding. We then present main features and performances of the proposed hybrid coder in Section 3 and 4, respectively. Finally conclusion is made in Section 5.

2. BACKGROUND

Recently, several methods [9-11] related to Harmonic-CELP hybrid coding which combines advantages of Harmonic and CELP coders at rates below 4 kbps have been presented. These coders are based on the linear predictive coding model as shown in Figure 1 and adopt the Harmonic coding algorithm for the voiced signal, and the CELP method for the unvoiced signal. In the hybrid

coders, the typical CELP for unvoiced coding is commonly used. And in [10][11], DFT for harmonic estimation and in [9] pitch detection algorithm with long delay is used. Therefore the coders require high computational complexity for both voiced and unvoiced coding.

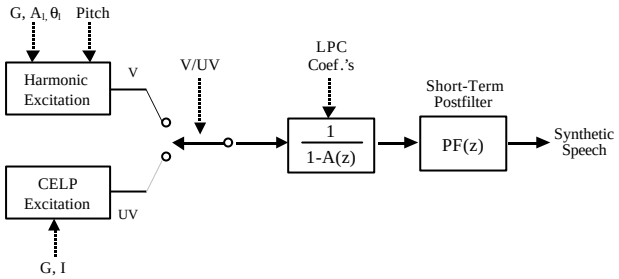

Figure 1. Simplified typical Harmonic-CELP hybrid model.

3. PROPOSED HARMONIC-CELP HYBRID CODER

In this paper, to overcome problems of the conventional hybrid coders, an Efficient Harmonic-CELP (EHC) hybrid coder shown in Figure 2 and Figure 3 is proposed, utilizing the well-known characteristics of the Harmonic and CELP type coders.

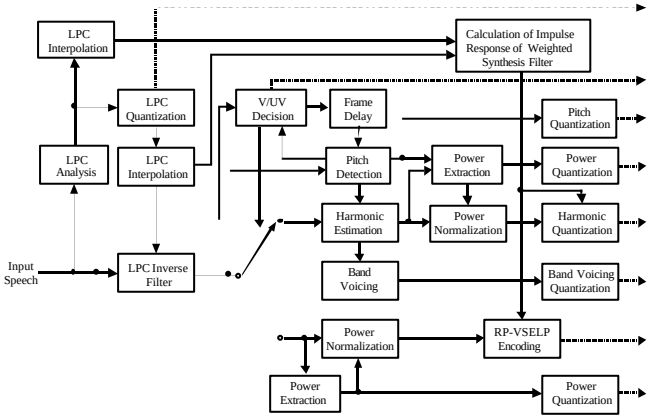

Figure 2. Proposed EHC hybrid encoder.

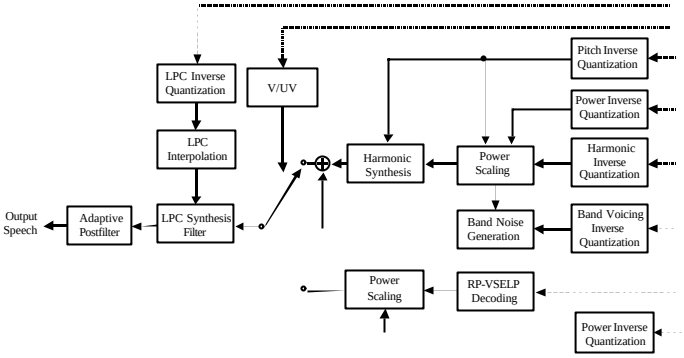

Figure 3. Proposed EHC hybrid decoder.

According to frame voicing decision, the proposed EHC hybrid coder employs a fast CELP algorithm for an unvoiced frame, and an improved Harmonic coding algorithm for an unvoiced frame. As a result, the proposed coder can produce good quality with low complexity.

3.1 LSP, Pitch and Voicing Decision

Line Spectral Pair (LSP) parameters are quantized via the Predictive Multi-stage Split Vector Quantization (PMSVQ) method similar to that in [12]. This technique is made up of three subsequent procedures: inter-frame prediction, full dimension VQ, and half dimension split VQ. Multiple LSP codebooks are separately designed and used for unvoiced and voiced speeches.

The modified pitch algorithm based on the simple forward tracking in [13] effectively avoids pitch doubling or halving by adopting the local maxima comparison in [12] and adaptively taking the first window in case of consecutive voiced frames, otherwise the second window as the reference.

A frame voicing value is firstly decided by observing signal and background noise levels, and then corrected using open-loop pitch gain, zero-crossing rate, and peakiness. Practically the EHC coder operates differently according to the previous and next frame voicing values.

3.2 Fast CELP Coding for Unvoiced Excitation

The well-known fast CELP algorithm such as the ACELP [8][12] using a few unit pulses is not appropriate for the unvoiced coding. In the proposed coder, the unvoiced excitation is encoded by means of the RP-VSELP [4], which has non-unit regular pulse basis vectors. The RP-VSELP significantly reduces the conventional VSELP [5] complexity while maintaining speech quality and robustness to channel errors. In addition, the RP-VSELP can optimize its codebook through an iterative close-loop training process [6]. Figure 4 shows a diagram of the RP-VSELP codebook search process. In this paper, to keep the continuous energy contour at very low bit rates, the signal gain is estimated

and quantized in an open-loop fashion. Therefore, only the signal shape is encoded using the RP-VSELP.

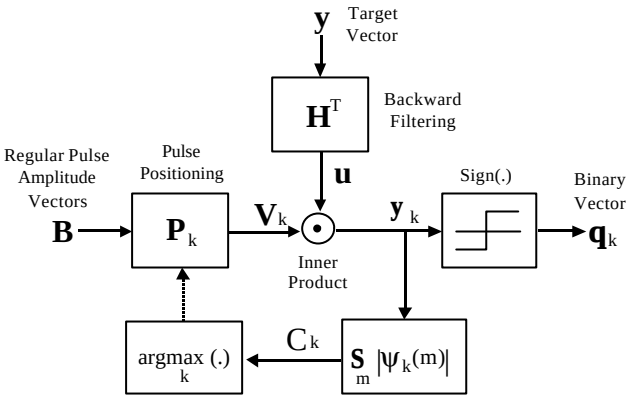

Figure 4. Block diagram of the RP-VSELP encoder.

3.3 Improved Harmonic Coding for Voiced Excitation

The improved Harmonic coding is proposed with the following features: fast harmonic estimation based on a Delta Adjustment (DA) method, Bark perceptual weighting for harmonic vector quantization and naturalness control by means of harmonic band voicing.

The DA method requiring only an integer pitch significantly reduces the complexity while maintaining the spectral distance performance, compared with the conventional harmonic estimation method [3][9] using a fractional pitch. Harmonic amplitude $A_l(\Delta_l)$ is estimated by minimizing the mean squared error $E_l(\Delta_l)$ between input spectrum $X_w(m+\Delta_l)$ and synthetic excitation spectrum $\hat{X}_w(m,\omega_0)$ using the following equations.

$$E_l(\Delta_l)=\sum_{m=a_l}^{b_l}\left[\left|X_w(m+\Delta_l)\right|-\left|\hat{X}_w(m,w_0)\right|\right]^2,\ -d_l\le\Delta_l\le d_l, \quad (1)$$

$$\hat{X}_w(m,\omega_0)=A_l\left|W(m,\omega_0)\right|, \quad (2)$$

$$\Delta_l=\left\lfloor\frac{\alpha\omega_0}{L-1}(l-1)\right\rfloor, \quad (3)$$

$$A_l(\Delta_l)=\frac{\sum_{m=a_l}^{b_l}\left|X_w(m+\Delta_l)\right|\left|W(m,\omega_0)\right|}{\sum_{m=a_l}^{b_l}\left|W(m,\omega_0)\right|^2}. \quad (4)$$

In (2), $W(m,\omega_0)$ is a window spectrum. Figure 5 shows a harmonic estimation process based on the proposed DA method.

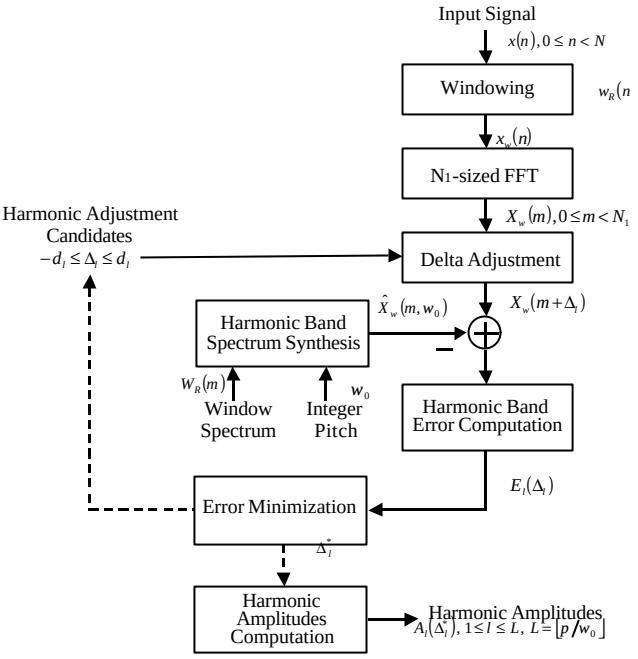

Figure 5. LPC residual spectrum and band voicing.

Complexity of the conventional method and the proposed method is on the average 13:1 from 4: 1 to 19:1. And spectral distortion between the conventional method and the proposed method is 0.1 dB that means the indistinguishable subjective quality degradation.

Estimated harmonic amplitudes are normalized and then converted with a fixed dimension by using the dimension conversion method [9]. Finally those are perceptually weighted vector quantized. This process is shown in Figure 6.

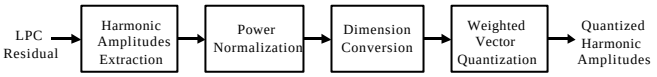

Figure 6. Harmonic amplitude quantization.

For harmonic vector quantization, new perceptual weighting is proposed. The proposed weighting function adds the psycoacoustic effect w_b derived from Bark frequency resolution $\Omega(f)$ [7] to the conventional perceptual weighting [2].

$$w_b=\left[\left.\frac{\partial\Omega(f)}{\partial f}\right|_{f=f_j}\right]^c\approx\left[\Omega(f)-\Omega(f_{j-1})\right]^c,\ 1\le j\le D, \quad (5)$$

$$\Omega(f)=6\log\left(\frac{f}{600}+\left(\left(\frac{f}{600}\right)^2+1\right)^{0.5}\right),\ 0\le f\le\frac{f_s}{2}, \quad (6)$$

In (5), $f_j = \frac{jf_s}{2D}$ and c controls the slope of the weighting curve.

At each harmonic band to control naturalness of the reproduced speech, band voicing values are computed by the following equation.

$$BV_l = 1 - E_l^{'} = 1 - \frac{\sum_{m=a_l}^{b_l} \left| |X_w(m+\Delta_l)| - |\hat{X}_w(m, w_0)| \right|^2}{\sum_{m=a_l}^{b_l} |X_w(m+\Delta_l)|^2} \qquad (7)$$

Figure 7 shows an example of the input LPC residual spectrum and the band voicing obtained by (7).

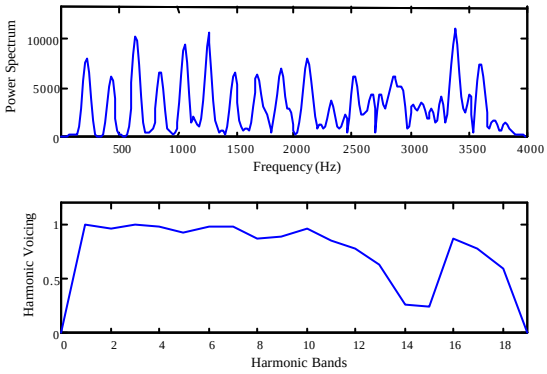

Figure 7. LPC residual spectrum and band voicing.

The band voicing vector is converted with a fixed dimension and then vector quantized.

In the decoder, the colored noise to be added to the synthetic harmonic excitation signal is generated by modifying a white gaussian noise spectrum using the harmonic amplitude and band voicing vectors.

The voiced excitation signal is synthesized by using the fast harmonic synthesis method based on over-sampling, time alignment and down-sampling [9].

4. EXPERIMENTAL RESULTS

To demonstrate performance of the proposed hybrid coding algorithm, we have implemented a 2.4 kbps EHC coder with a bit allocation shown in Table 1. The implemented coder has 25 ms frame (200 samples at 8 kHz), 10 ms lookahead , 10 ms lookback. LPC parameters that are analyzed using 25 ms hamming window are converted to LSP and weighted vector quantized using separate codebooks, according to frame voicing. For an unvoiced frame, the frame is divided into four subframes. At each subframe, an open-loop LPC excitation power is 4-bit non-uniform scalar quantized, and the excitation shape is vector quantized using a 5-bit RP-VSELP codebook. For a voiced frame, pitch is 6 bit non-uniform scalar quantized. Harmonic excitation power is 5-bit scalar quantized in an open-loop fashion, and harmonic amplitude vector is 5-bit vector quantized twice a frame. But band voicing vector is 3-bit vector quantized once a frame.

Table 1. Bit allocation.

	Unvoiced	Voiced
LPC	23	24
V/UV	1	1
Pitch		6
Power	4×4	5×2
Shape	5×4	8×2
Band Voicing		3
Total Bits / 25ms	60	60

Subjective MOS (Mean Opinion Score) test was performed and its results is shown in Table 2. Form the results, we can see that the proposed 2.4 kbps EHC coder has good quality comparable to that of the 4.4 kbps IMBE.

Table 2. Results of the MOS test.

	5.3 kbps ACELP	4.4 kbps IMBE	2.4 kbps EHC
Female	3.35	3.20	3.16
Male	3.65	3.15	3.23

5. CONCLUSION

An efficient Harmonic-CELP hybrid coding algorithm utilizing the well-known characteristics of the Harmonic and CELP coders has been proposed. The proposed hybrid coder employs the RP-VSELP coder as a fast CELP in case of unvoiced, or an improved Harmonic coder in case of voiced. Features of the new coder are simple pitch detection, fast harmonic estimation, variable dimension harmonic vector quantization, perceptual weighting including Bark frequency resolution, fast harmonic synthesis, and naturalness control by harmonic band voicing. From results of subjective tests, the proposed hybrid coder demonstrated good quality with reasonable complexity at about half rates of the reference coders.

6. REFERENCES

[1] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp.25.1.1-25.1.4, 1985.

[2] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995.

[3] *IMBE Vocoder Description*, DVSI, Jul. 1993.

[4] Y. S. Choi, S. W. Park, and D. H. Youn, "Fast Vector-Sum Codebook Search Method for Low Bit Rate Speech Coding," *IEE Electronic Letters*, Vol. 33, No. 6, 1997.

[5] I. A. Gerson and M. A. Jasuik, "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kbps," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp.461-464, 1990.

[6] Y. S. Choi, H. G. Kang, and D. H. Youn, "A Fast VSELP Speech Coder Based on Mutually Orthonormal Regular Pulse Vectors," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp.554-557, 1996.

[7] E. Zwicker and H. Fastle, *Psycoacoustics: Facts and Models*, Springer-Verlag, 1990.

[8] ITU-T Recommendation G.723.1, *General Aspects of Digital Transmission Systems: Dual Rate Speech Coder for Multimedia Communications Transmission at 5.3 and 6.3 kbit/s*, 1996.

[9] *Technical Description of Sony IPC's Proposals for MPEG-4 Audio and Speech Coding*, Sony, Nov. 1995.

[10] W. B. Kleijn, "Encoding Speech Using Prototype Waveforms", *IEEE Trans. on Acoust. Speech and Signal Proc.*, Vol. 1, No. 4, pp.386-399, 1993.

[11] J. C. De Martin and A. Gersho, "Mixed-Domain Coding of Speech At 3 kbps," *IEEE Proc. Int. Conf. Acoust., Speech, Signal Proc.*, pp.216-219, 1996.

[12] ITU-T Recommendation G.729, *Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic Code Excited Linear Prediction (CS-ACELP)*, 1995.

[13] Proposed TIA/EIA/PN-3292 Standard - *Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*, Official Ballot Version, Qualcomn Inc., Apr. 1996.