ISCA Archive
http://www.isca-speech.org/archive

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

A New Low Bit Rate Speech Coder Based On Intraframe Waveform Interpolation

Hongtao Hu and Du Limin
Institute of Acoustics, Chinese Academy of Sciences
Beijing, P.O. BOX 2712, 100080, P.R. China

ABSTRACT

A new characteristic waveform (CW) interpolation coder is proposed in this paper. In the proposed coder, two characteristic waveforms are extracted from LPC residual signal at each frame. The Waveform Interpolation (WI) is operated within the frame. In the novel WI, variable dimension vector quantization (VDVQ) and power vector quantization are proposed and the low frequency band (LFB) and high frequency band (HFB) are allocated different numbers of bits according to human hearing perception. A result of a 2400 bps coder is presented and the reconstructed speech shows its quality is closed to FED-STD-1016.

1 INTRODUCTION

Many telecommunication standard coders at 4 to 8 kbps use the code excited linear predictive (CELP) [1] coding algorithm based on analysis-by-synthesis method and it can produce communication quality speech signals. When the bit rate becomes less than 4 kbps, however, its quality will be damaged because the model cannot produce natural voiced speech.

Several coders have been proposed at bit rates below 4 kbps. There are 3 main algorithms: Sinusoidal Transform Coder (STC) [2], Multiband Excitation Coder (MBE) [3], and Waveform Interpolation Coder (WIC) which was pioneered by Doc. Kleign [4-5]. WIC has the potential to produce toll quality recovered speech at low bit rate below 4 kbps. WI allows the efficient compression of speech signals by exploiting human perception to speech signals. The speech signal is represented by an evolving characteristic waveform and then decomposed into a slowing evolving waveform characterizing voiced speech and a rapid evolving waveform characterizing noise-like speech [5].

The decomposition results two fundamentally different waveform components and different quantization is required. The decomposition is motivated by human perception and allows efficient coding at bit rate between 2 and 4 kbps. However, REW magnitude surface is not pure and contains significant components of the SEW magnitude [6]. This highlighted the necessity for an improved decomposition method [7]. In this paper, a different WI method is proposed. The LPC residual signal is represented in terms of pitch synchronous transform. Two character waveforms are extracted in each frame. Waveform interpolation is performed directly within the frame and no decomposition of transmission frequency band is adopted. The characteristic waveform is quantized through VDVQ and the characteristic waveforms are divided into two frequency bands: high frequency band and low frequency band. More bits are allocated to low frequency band than high frequency band.

2 SPEECH MODEL

The conventional WI vocoder performs good quality, however, its computational complexity is very high and is delay is long [4]. The complexity results from 1-dimentison to 2-dimention representation of speech signal. A new 1-D to 2-D representation is adopted in this paper and he intraframe interpolation is feasible. The complexity is reduced and the delay is also shortened.

The new speech model is pitch-synchronous representation of speech [10]. We assume that a sequence P(k) of integer local pitch periods has been extracted from a sampled speech signal s(n). We can store each period-length segment of signal s(n) in a vector $r(k)=[r_0(k),r_1(k),\ldots,r_{P(k)-1}(k)]$, obtaining a 2-D representation of the 1-D speech signal.

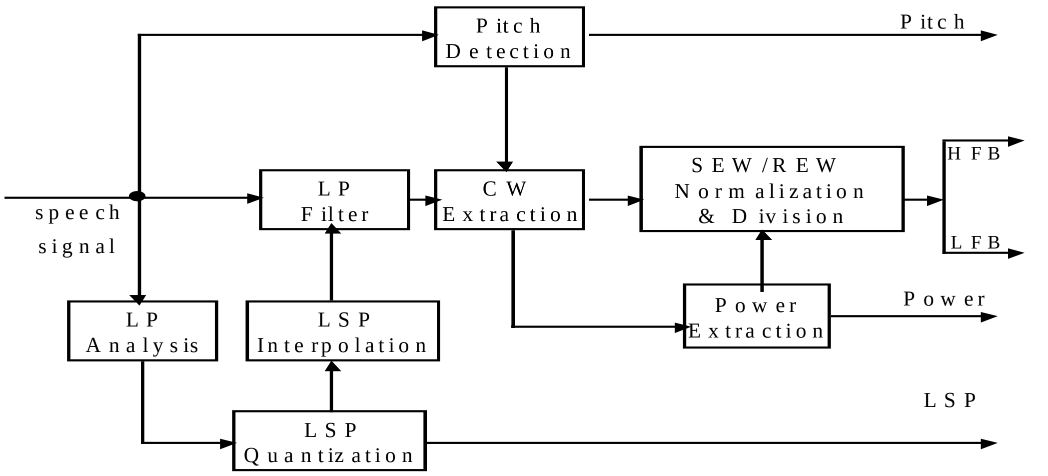

Fig1 Encoder Block Scheme

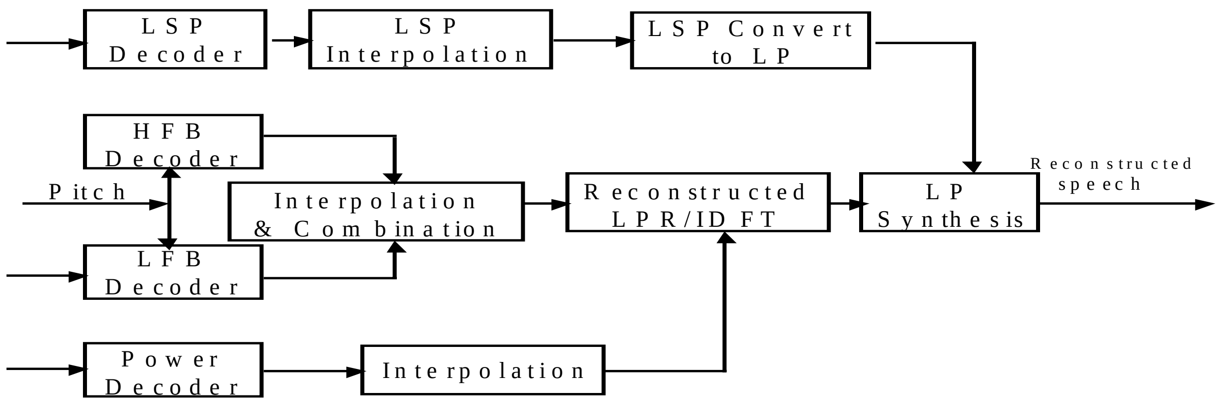

Fig2 Decoder Block Scheme

Several waveform interpolation algorithm based on pitch synchronization are proposed [8,9]. But most of them are partial coders. They can produce efficient compression of voiced speech but no efficient compression of unvoiced speech. So some other methods are combined to process the unvoiced speech and U/V determination is needed. This may result in discontinuity during the switches. In this paper, a novel coders which no other coding methods are added is proposed. Two characteristic waveform are extracted each frame and the interpolation is operated within the frame.

3 CODING SECHEM

In fig. 1 and fig. 2, the block diagrams of the proposed 2400 bps coder and the decoder are reported. The coder extracts characteristic waveforms from LPC residual signal. The characteristic waveforms are transformed to frequency domain and normalized then divided into two bands.

4 CHARACTERISTIC WAVEFORM EXTRACTION

The waveform interpolate is performed in LPC residual domain. The LPC residual signal is to be split into a sequence of characteristic waveform p_I ($0 \le I \le M-1$). The M_n characteristic waveforms in the nth frame form the matrix R_n.

$R_n=(p_0, p_1, \dots p_{M_n-1})$ (1)

$p_I=(r(I,0),r(I,1), \dots r(I,N_n-1))^T$

$0 \le I \le M_n-1$ (2)

Where M_n is the number of pitch waveforms in the frame whose length is L_f. In the following, we ignore the subscript n. N is the pitch period.

$$M = \begin{cases} \dfrac{L_f}{N} & if\ L_f \bmod N = 0 \\ \left\lfloor \dfrac{L_f}{N} \right\rfloor + 1 & if\ L_f \bmod N \neq 0 \end{cases} \qquad (3)$$

Fig. 3 shows the extracting characteristic waveform procedure in one frame. The first characteristic waveform contains some samples of the

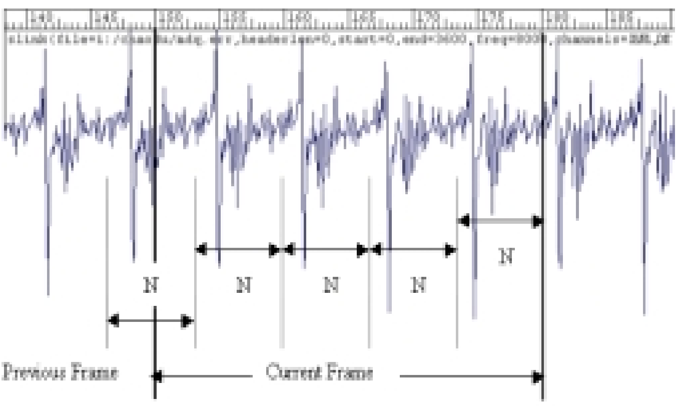

Fig3

previous frame. The characteristic waveforms are transformed to frequency domain and normalized to have a magnitude of unity.

5 QUANTIZATION OF CW AND POWER

The dimension of CW spectra is variable. The variable dimension vector quantization (VDVQ) is a challenging question. An efficient method is proposed [11]. But it is for the vocoder of MBE. The following is the new VDVQ used in this new coder. We use a universal codebook whose dimension is fixed. Any input vector S whose dimension is I can be mapped into a vector X from the universal codebook whose dimension is K ($I \le K$). The following is the conversion.

(a) From I dimension to the fixed dimension:

$$\begin{cases} X(k) = S(i), k = \left\lceil \dfrac{K \times i}{I} \right\rceil, 0 \le i \le I \\ 0.0, otherwise \end{cases} \qquad (4)$$

$0 \le k \le K$

(b) From the fixed dimension to the I dimension

$$S(i) = X(k), k = \left\lceil \dfrac{K \times i}{I} \right\rceil \qquad (5)$$

$0 \le k \le K$

The powers of the two CW in the same frame are extracted and formed a vector. Then the power is vector quantized.

The complete bit allocation is reported in Table 1.

Table 1 Bit allocation for the 2400 bps WI coder

Parameter	bits/Frame
LSF	24
Power	10
Pitch	7
CW	2*(9+6)

6 SIMULATION RESULTS

Twenty sentences (ten male sentences and ten female sentences respectively) were tested. All the tests show that the reconstructed quality is close to FS1016. Figure 4 shows an example of the vocoder. The top is the original speech. The middle is reconstructed speech but the CWs are unquantized but the other parameters are quantized. The bottom is the full reconstructed speech.

References

[1] M. R. Schroeder and B. S. Atal, " Code-Excited Linear Predictive (CELP): High Quality Speech at Very Low Bit Rate," Proc. ICASSP-85, pp. 937-940, 1985.

[2] R.J.McAulary and T.F.Quatieri, “ Speech analysis /synthesis based on a sinusoidal representation,” IEEE Trans. Acoustic Speech Signal Processing, vol.34,pp.744-754, 1986.

[3] D.W.Griffin and J.S.Lim, “Multiband excitation vocoder,” IEEE Trans. Acoustic Speech Signal Processing, vol.36,no.8, pp.1223-1235, 1988.

[4] K.B.Kleijn and K.K.Paliwal, eds., Speech Coding and Synthesis. Elsevier, 1995.

[5] W.B.Kleijn and J.Haagen, "Transformation and decomposition of the speech signal for coding," IEEE Signal Processing Letter, vol.1, no.9, pp.136-138, 1994.

[6] N.R.Chong, I.S.Burnett, J.F.Chicharo, M.M.Thomson, "The effects of noise on the waveform interpolation speech coder," Proceedings of IEEE TENCON'97, Brisbane, Australia, Dec,,1997.

[7] N.R.Chong, I.S.Burnett, J.F.Chicharo, M.M.Thomson, "Use of the Pitch Synchronous Wavelet Transform as a New Decomposition Method for WI," in Proc. Int. Conf. Acoustic Speech Signal Processing, pp.I513-I516, 1998.

[8] Michele FESTA and Daniele SERENO, "A speech coding algorithm based on prototypes interpolation with critical bands and phase coding", in Proc. Of Eurospeech'95, Madrid, Sept. 1995, pp.229-232.

[9] Y.Tanaka and H.Kimura, "Low-bit-rate speech coding using a two-dimensional transform of residual signals and waveform interpolation," in Proc. Int. Conf. Acoustic Speech Signal Processing, pp.I173-I176, 1994.

[10] Gianpaolo, Evangelista, "Pitch-Synchronous Wavelet Representations of Speech and Music Signals", IEEE Tran. on Signal Processing, vol.41,No.12,1993.

[11] A.Das and A.Gersho, "Variable Dimension Vector Quantization of Speech Spectra for Low-Rate Vocoders", IEEE Proc. Of the Data Commpression Conference, pp.420-429, April 1994.

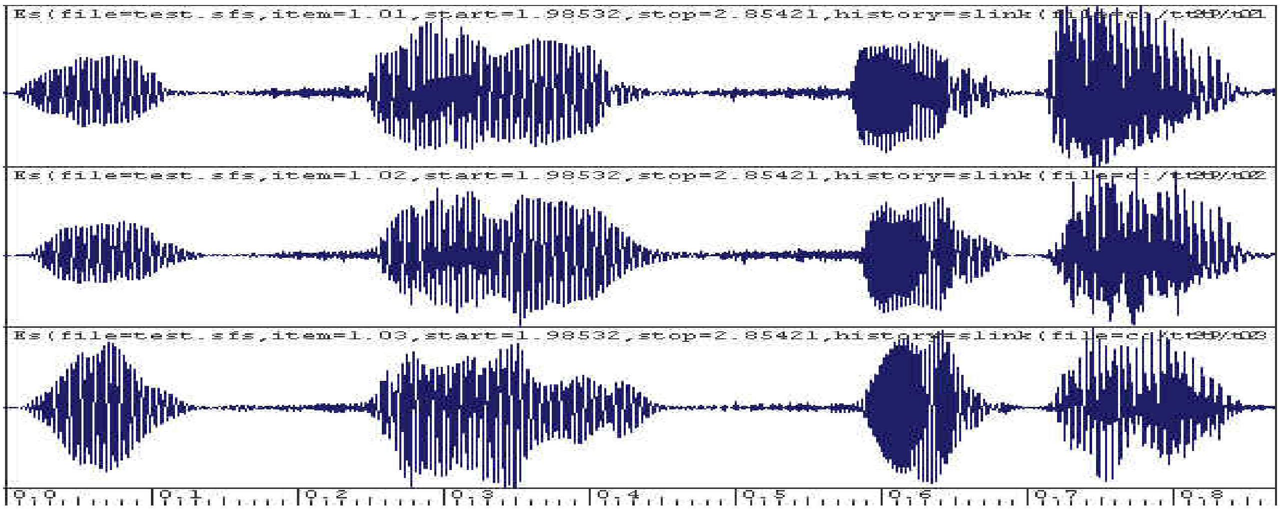

Fig.4 An example of the results