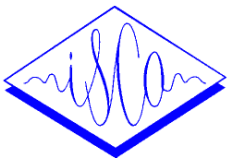


6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

A MIXED AND CODE EXCITATION LPC VOCODER AT 1.76KB/S

Jinzhong Zhang and Yingmin He and Renshu Yu

Electronic Engineering Department, Harbin Engineering University, Harbin, China

ABSTRACT

This paper describes a speech-coding algorithm that we develop recently. We call it M&CELP VOCODER. It is based on the traditional linear prediction vocoder. The excitation is constructed by not usually one of either periodic pulse or white noise, but the mixture of periodic feature-waveform (from predict error) and periodic pulse and white noise passed high pass filter in voiced speech section and white noise in the unvoiced speech. At parameter quantization, The quantizer based on Tree-search technique is applied, and it is improved. It enhanced the performance of reconstructed speech, while maintaining the lower bit rate. In addition, the scheme of pitch evaluation is improved, and this vocoder is very efficient to eliminate thumps and buzz, and it enhanced the naturalness and intelligibility. The algorithm is very efficient at speed, it can be realized at the PC computer with high configuration. Informal listening confirm that it is as good as the new U.S. federal standard: MELP at 2.4kb/s.

1. INTRODUCTION

The task of speech coding is to reduce the bit rate of signal source. Simultaneous, it is necessary to remain the good performance of perception. The motivation of speech coding is to reduce the storage of speech and the bandwidth occupied by speech. At present, many algorithms about speech coding have been developed. For example: LPC-10, MPE-LPC RPE-LPC and CELP VOCODER. But they are not very perfect. If bit rate reduces to below 2.4kbps, it is very difficult that the quality of synthesized speech is accepted. Perhaps, you can understand the meaning of the speaker, but the naturalness and intelligibility is not ideal. It is more difficult that the speaker is identified. The other fault of these coders is complicated in the structure. This vocoder reduces the complexity of coder and maintains the perfect quality of synthesized speech.

2. THE SPECIFICATIONS

2.1. The Frame Size of Speech

The frame size taken by this algorithm is 200 samples (25ms and the sample is 8kHz) that is quit fit for most speech signal, The frame length of the LPC analysis is 220 samples. Because the length of pitch period arranges from 20 to 120 samples(for 8 kHz sample rate) generally, the size of pitch window is 240 samples.

2.2. The Demand for Analog Speech Signal

Because this vocoder processes the speech signal whose the bandwidth is limited to telephone, the speech signal must be passed a 300-3800Hz band-pass filter in advance before it is sampled with 8000Hz sample rate and quantized with 8-bit quantizer.

3. THE DESCRIPTION OF ALGORITHM

3.1. Introduction about Coder

This M&CELP vocoder is based on usual LPC vocoder model. But it is improved in many parts; for instance, many new features, the mixed excitation, adaptive spectral enhancement, pulse dispersion filter and feature-waveform are applied. The approach of evaluation of pitch period is more simple and faster. In addition, the quantizer is also improved. Block diagram 1 shows the block diagram of the coder.

In this paper, the excitation signal is constructed as below, the feature-waveform from residual signal is repeated periodically by pitch period in advance, then the white noise passed high pass filter [3] is added to it. So the reconstructed excitation is more consistent with the original residual signal. This method reduces the thumps and buzzes in the synthesized speech and matches the synthesed speech signal to original speech signal better with auxiliary technique.

3.2. Coder

3.2.1. Remove Low Frequency

Above all, we pass digitized speech signal a high-pass filter to remove low frequency interference. This filter is a Chebyshev II type one with 60Hz cut frequency and 30db attenuation.

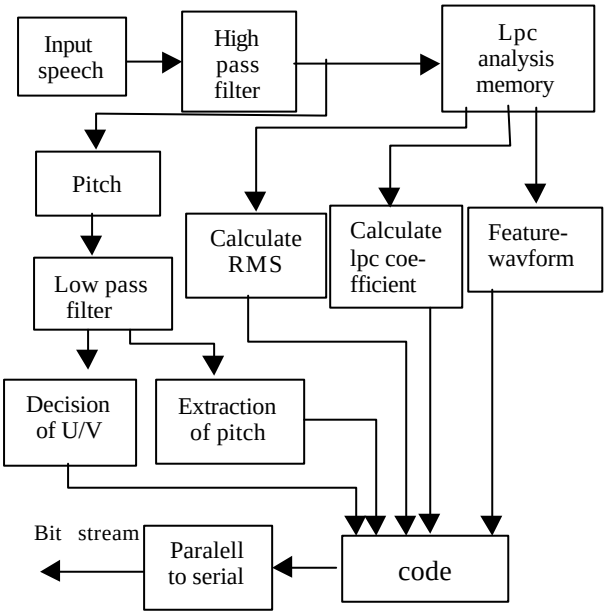
3.2.2. The Decision of U/V

As demonstrated by other voicing algorithms, linear discrimination is appropriate for voicing classification. For the two-class voiced/unvoiced problem, using N parameters, the M-level adaptive linear discriminant classifier is [7]:

$$\sum_{i=1}^{N} a_{i,j} p_i + c_j > 0, \text{where } j \in \{0, \mathrm{K}, \mathrm{M}-1\}$$

And where $a_{i,j}$ and c_j are weights and p_i is the measured signal parameter. The discriminator classifies the speech segment as voiced if the expression is true; otherwise, classify it as unvoiced. The $a_{i,j}$'s and c_j's adapt to an acoustic noise level by selecting j according to the signal-to-noise ratio (SNR).

One technique to determine the weights $a_{i,j}$ and c_j is fisher's Optimal Linear Discriminant Analysis method. It is optimal under multi-variate-normal and equal covariance parameter sets. Fisher's method chooses the coefficients that maximize the ratio of the difference of the means of the linear combination in the two groups to their common variance. Although this method assumes equal covariance, in our case, the matrices are close enough that it makes little or no difference in the results to assume equality. In addition, this method is quite robust to non-normality.



Block diagram 1: The block diagram of coder

3.2.3. The Calculation of Pitch

Above all, the autocorrelation function of speech has been calculated. Because the autocorelation function is also periodic signal, and the arrange of pitch period of speech is from 20 to 120 samples, the pitch period is the time lag corresponding to the maximum of autocorelation function between 20 and 120. But this method is easily interfered by the formant, so this results in error. In this matter, it results in reducing the quality of the speech, so we must eliminate the influence. Where we propose a new clip waveform method.

1. Divide a frame speech into 4 subframes.
2. Find the absolute value of each sample in each subframe.
3. Find average absolute value of each subframe.
4. Subtract 1.35 average absolute value from each sample in subframe.(where 1.35 is experimental value)
5. Determine pitch period according to autocorrelation.

Given s (n) is a section of speech with window, it's no-zero interval is $n = 0 \sim (N-1)$, where N is the length of frame . The autocorrelation function is expressed as:

$$R(l) = \sum_{n=-\infty}^{\infty} s(n)s(n+l) = \sum_{n=0}^{N-l-1} s(n)s(n+l)$$

3.2.4. Extraction of the Feature-waveform

Literature [4][5] proposed that the first ten harmonic magnitude of the Fourier Transform of the residual signal be transmitted. When synthesing, the first ten harmonic magnitude is transformed the pitch period samples in terms of DFT, then it is repeated periodically by the pitch period. Literature[6] proposed that the residual signal of a pith period be extracted from original residual signal. When synthesing, it is also repeated periodically by the pitch period. Because the period is variable, it is very difficult that the parameter is quantized. The calculation quantity of this method is very large.

After linear predict coefficients are determined, we can pass speech signal the inverse filter to find residual signal. Figure 1 shows the residual signal of a voiced speech. From figure 1, the energy of residual signal is very large in the residual signal correspond to glottal pulse, and the other is very small, so it can be looked as white noise.

In this paper, we extract only 30 samples in the part having larger amplitude and call it feature-waveform. In addition, the residual signal is quasi-period in a frame, so the only feature-waveform is transmitted, and when decoding, it is repeated periodically according to the pitch period.

Figure 1: Original speech and Residual signal

3.2.5. Calculate Energy

For calculation of energy, we use RMS:

$$RMS = \sqrt{\frac{1}{M}\sum_{i=1}^{M} S_i^2}$$

3.2.6. Quantization Of Parameter

1. Quantization of LPC Coefficient

Above all, The LPC coefficients are converted to LSF's[1][2]. The division vector quatization method is applied. Namely, 10 LSF's is divided into two groups first four as a group and last six as a group, then quantizing them respectively. We adopt tree-search vector quantization. We find that tree-search can be improved by experience. When sample set is deviled into some

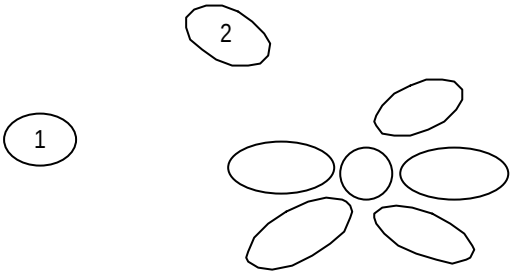


Figure 2: cell distribution

level, some cell may contain little samples. Even it contains only 1 to 2 samples, figure 2 shows this matter. For 1, it can be looked as interference, we eliminate it. For 2 we think the probability of it is very small, so assign it a codeword and do not devoid it, the other cells continue to be deviled. The rest may be deduced by analogy. The VQ method makes the search and training of codebook more accurate and faster. In another words, the samples with larger probability are quantized accurately and the ones with little probability are quantized coarsely. We call it non-even VQ.

2. Quantization of Remaining Parameter

To reduce bit rate, for LSF's and feature-waveform, we also use vector quantization. For pitch period and energy, we use uniform quantization. Table 1 shows bit allocation:

	Voiced speech	Unvoiced speech
LSF's	12+12	12+12
Energy	5	5
pitch	6	0
Feature-waveform	8	0
U/V	1	1

Table 1: Bit allocation

The total bit is 44 bit's, so bit rate is 1.76kbps.

3.3. Decoder

The decoding is inverse procedure of the coding, diagram 2 shows docoding flow procedure. When decoding, LSF's are converted to LPC coefficient, and the excitation is reconstructed by adding white noise to feature-waveform repeated periodically by pitch period.

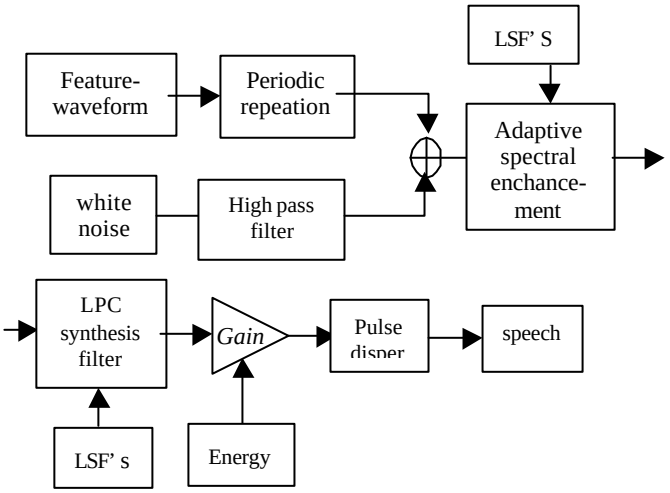
3.3.1. Adaptive Spectral Enhancement

Due to the limitation of all poles model, it is not too consistent with spectrum of the original speech signal, especially at the peak and valley of the spectrum. The adaptive spectral enhancement takes advantage of the pole of linear predict filter. It can enhance the peak and valley of spectrum of speech signal, so it can match synthesized speech with original speech in the peak and valley of spectrum of speech signal.

The excitation reconstructed need pass an adaptive spectral enhance filter[4][5] in advance, this filter comprise of a tenth order pole and zero filter and a first order tilt filter. Its coefficients are LPC ones with bandwidth expansion. And it's transmission function as follow:

$$H_{ase}(z) = \frac{A(\alpha \cdot z^{-1})}{A(\beta \cdot z^{-1})}(1 + m \cdot z^{-1})$$

Where $\alpha = 0.5, \beta = 0.8, m = \max(0.5k_1, 0)$, k_1 is the first PARCOR(part correlation coefficent).



Block diagram 2: Decoder block diagram

3.3.2. Pulse dispersion

After the adaptive spectral enhancement, the output signal is passed into the LPC synthesed filter. Then it is scaled by

energy. In the end, to implement pulse dispersion, we let synthesed speech signal pass a triangle pulse filter based on spectral flat. This can let speech signal become more flat in entire period.

4. CONCLUSION

Using this algorithm, speech reconstructed sound very intelligible and natural. This paper proposed a new clip waveform method and a new constructed excitation method and a non-even VQ technique to efficiently encode the LPC and waveform. Listening tests show that this coder performs as well as the new U.S. federal standard: MELP at 2.4kb/s.

5. REFERENCES

1. Peter Kabal: The Computation of line spectral frequencies using Chebyshev polynomials 1986 IEEE ICASSP.
2. A.Grassi,A.Dufaux: Efficient algorithm to compute LSP parameters from 10th -order LPC coefficients, 1997 IEEE ICASSP.
3. Alan V. McCree and Thomas P. Barnwell: A new mixed excitation LPC vocoder 1991 IEEE ICASSP
4. Alan McCree and KwanTruong: A 2.4kbps MELP coder candidate for the new U.S. federal standard 1996 IEEE ICASSP
5. Alan McCree: A 1.7kb/s MELP coder with improved analysis and quantization 1997 IEEE ICASSP
6. Yuusuke Hiwasaki and Kazunori Mano: A new 2-kbit/s speech coder based on normalized pitch waveform 1997 IEEE ICASSP
7. Joseph P. Campbell: voiced/unvoiced classification of speech with applications to the U.S. Government LPC-10e algorithm 1986 IEEE ICASSP
8. XingJun Yang *Speech signal processing* publishing house of electronics industry