The Pitch Extraction Method through Spectrum Flattening

SOYEON MIN*, KYUNGA JANG, MYUNGJIN BAE****

Department of *Electronics Engineering
Department of ** Information and Telecommunication Engineering
Soongsil University,
1-1 Sando 5dong Dongjakgu, Seoul, KOREA, 156-743

Abstract

The exact pitch(fundamental frequency) extraction is important in speech signal processing like speech recognition, speech analysis and synthesis. However the exact pitch extraction from speech signal is very difficult due to the effect of formant and transitional amplitude. So in this paper, the pitch is detected after the elimination of formant ingredients by flattening the spectrum in frequency region. The effect of the transition and change of phoneme is low in frequency region. In this paper we proposed the new flattening method of log spectrum and the performance was compared with LPC method and Cepstrum method. The results show the proposed method is better than conventional method.

Key-words: Pitch Extraction, Flattening Method, Autocorrelation method

1 Introduction

Fundamental frequency in speech signal processing field, pitch information is very important. If fundamental frequency of speech signal can be detected well, the accuracy of recognition also can be higher due to a effect decrement of speaker in speech recognition and be changed easily or maintain the natural and characteristic in speech synthesis. Also, the effect of glottis can be removed and get the parameter of correct vocal track if the pitch synchronized is analyzed. Because of this importance of the pitch detection, methods about pitch detection have been proposed variously and it can be divided by time, frequency and time - frequency domain method. The time domain detection method is simple. There are parallel processing, AMDF and ACM method etc but the pitch detection is very difficult in transition region. Pitch detection methods of frequency domain are the harmonic analysis method [1,3], Lifter method and Comb-filtering method. This method isn't given a effect by the change or transition of phoneme but if the point number of FFT(Fast Fourier Transform) is increased in order to increase the detection of fundamental frequency the processing time is longer as much as the point number is increased and it's insensitive to the change of characteristic. Time - frequency domain method takes a advantage of the time domain method and the frequency domain method. This technique have Cepstrum method, spectrum comparison method etc. and apply the time and frequency domain both so the computational process is complicated as a disadvantage[3,4]. This paper proposed the accurate pitch detection that the effect of formant can be removed by flattening the spectrum and the resolution of frequency also can be increased without increasing the number of FFT point. The spectrum flattening technique and the pitch detection used its technique describes in section 2 and 3. In section 4, experiment and result describes and conclude in section 5.

2 Spectrum Flattening Process

Speech signal transform FFT (Fast Fourier Transform) to the frequency domain and the analysis of spectrum is done in the frequency domain. Figure 1 is a block diagram of spectrum flattening algorithm proposed in this paper. Divide the frequency band by some sub-band as the first step to remove the effect of the formant and of the transition amplitude from spectrum signal. At this time, the bandwidth of sub-band causes much effect to spectrum flattening. Pitch period is about 2.5-25msec so the bandwidth of sub-

band takes 300Hz and 400Hz. This is for progressing adaptively depending on input speech. Next step, maximum value in each sub-band stores as a parameter of frame. The values of parameter are about 10-13 with 8kHz sampling rate. Those values can do the modeling of formant envelope well because it reflects formant component directly. After linear interpolation by parameters and the obtainment of formant envelope approximately, we extracts its formant envelop from spectrum signal. This is the first spectrum flattening. Most ideal result can be obtained when the sub-band width is decided by the pitch period of input speech. Therefore, the second spectrum flattening is progressed with the signal flattened once via above algorithm again in order to compensate the result of the first spectrum flattening. This time, sub-band's bandwidth used bandwidth of each 3 case. When the bandwidth of the first flattening was 300Hz and 400Hz, we used 300Hz, 400Hz, 500Hz and 200Hz, 300Hz, 400Hz bandwidth each. Comparison estimation method about each result used variance. Before calculating the variance, each result signals does the normalization for making the maximum value being 0(zero). Variance used in this paper is as following.

$$Variance = \frac{2}{N}\sum_{k=1}^{N/2}(x(k)-m)^2 \qquad (1)$$

Where N is the number of FFT point and spectrum signal is symmetry by Y axis so the variance progress to N/2. Also, k is sample index in frequency domain and *m* is mean average. *m* value is evaluated the standard of flattening as a standard with 0(zero). Figure 2 is the speech waveform in voiced region and log-spectrum signal. Figure 3 shows the result of flattening the speech signal by using the proposed algorithm. (a) and (b) of Figure 3 are a example of flattening the spectrum by using LPC(Linear Prediction Coding) as representative formant modelling method and Cepstrum method. Proposed flattening techniques has better result like figure shows. Figure 4 is a signal in the transition region and its log-spectrum and figure 5 is the result of flattening this signal. Similarly, we can know that the performance of the spectrum flattening is better than LPC or Cepstrum method in transition region. Because of this excellent performance, the pitch can detected accurately in transition region as well as voiced region.

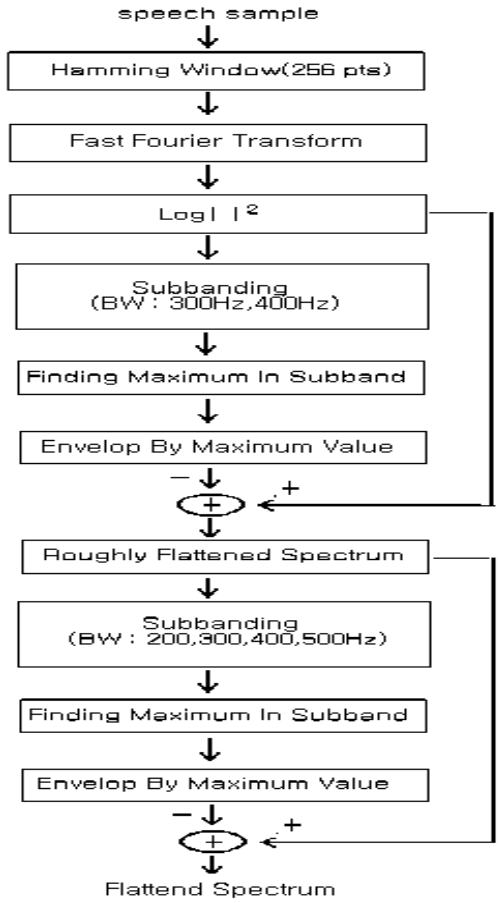

Fig. 1 The Flattening Process of Spectrum

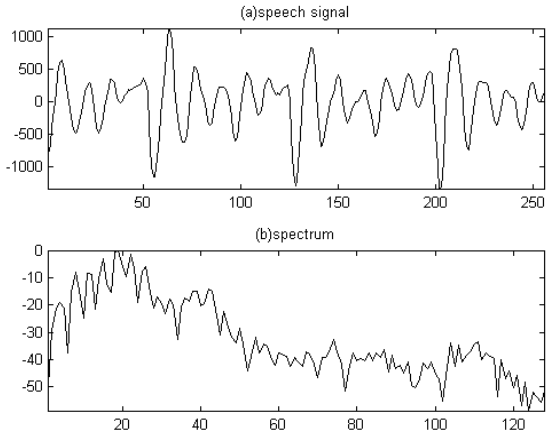

Fig. 2 Signal in Voiced Region
(a) Time Domain Signal
(b)Log Spectrum Signal

3 Pitch Detection Process

Used autocorrelation method to get fundamental frequency(pitch) from flattened spectrum signal. When P(k) is log-spectrum signal, autocorrelation method is defined as following. M display a number of sample delayed in frequency domain here.

$$R(m) = \frac{2}{N}\sum_{k=2}^{N/2} P(k)P(k+m) \qquad (2)$$

P (k) may achieve 1/2 of N that autocorrelation is FFT size because is left and right symmetry. As pass through pre-emphasis process to use effective autocorrelation method in time domain, need pre-emphasis process similarly in frequency domain. First, autocorrelation method should be applied in stability section. However, do analysis period to limit wide-band to 0 - 1 [kHz] because harmonics that is not stable in high frequency domain appears. Also, must consider frequency resolution. Frequency resolution is proportional in the number of FFT points, but the length is limited always. Therefore, do sign linear interpolation to compensate frequency resolution. This can do more correct pitch detection.

4 Experimental Results

Computer simulation was performed to evaluate the proposed algorithm using an IBM Pentium III interfaced with the 16-bit AD/DA converter. To measure the performance of the proposed algorithm, we used the following speech data. Speech data was sampled at 8kHz and was quantized 16bits. Following sentences were uttered five times by 5 male and female speakers who are in the middle or later twenties. The data were recorded in a quiet room, with the SNR(Signal to Noise Ratio) greater than 30dB.

Sentence 1: /Insune komaneun cheonjaesonyuneul joahanda/
Sentence 2: /Yesunimkeoseo cheonjichangjoeu kyohuneul malseumhasyuda./
Sentence 3: /Changgongeul hechye naganeu inganeu dojeoneun keuchieobda
Sentence 4: /Soongsildaehakgyo eumseiongtongshin yeungusilida/

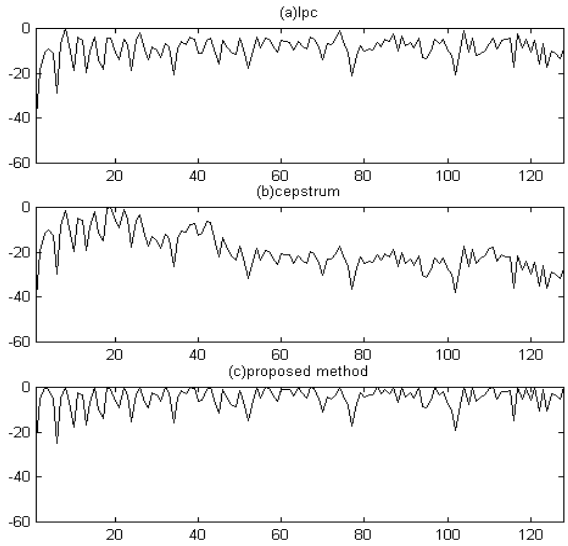

Fig. 3 Flattened Spectrum Signal
(a) LPC Method (b) Cepstrum Method
(c)Proposed Method

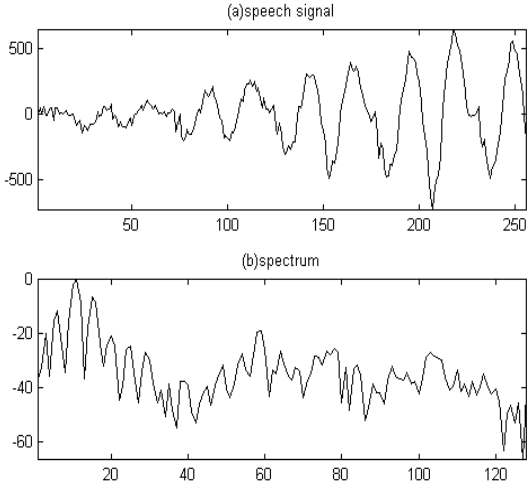

Fig. 4 Signal in Transitional Region
(a) Time Domain Signal
(b) Log Spectrum signal

Table 1 shows variance in case of male speaker.
As we can be seen in Table 1, the Cepstrum method shows the large variance and LPC method shows good characteristic. But, in case of LPC method, we get large variance about 1.5 times than proposed algorithm.

Figure 7 is pitch contour in case of using utterance 1 in SNR 30dB environment. In figure 7, the experimental results show the proposed algorithm is better than LPC method and Cepstrum in the side of exactly pitch detection. Figure 8 is experimental result in SNR 6dB environment. Similary, the experimental results show the proposed method is better than LPC method and Cepstrum.

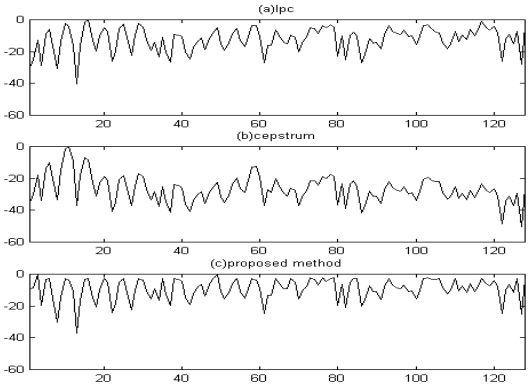

Fig. 5 Flattened Spectrum Signal
(a) LPC Method (b) Cepstrum Method
(c) Proposed Method

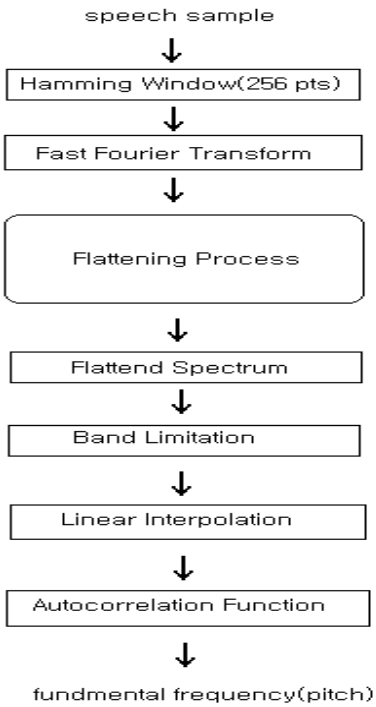

Fig. 6 The Process of Pitch Detection

Table 1. The Variance of Male Speaker[dB]

	LPC	Cepstrum	Proposed Method
Sentence1	187.13	716.23	124.04
Sentence2	169.02	697.20	108.68
Sentence3	179.12	704.65	119.20
Sentence4	163.72	680.17	107.97
Average	174.74	699.56	114.97

Fig. 7 Pitch Contour(30dB)
(a)Reference Pitch (b) LPC Method
(b) Cepstrum Method (d) Proposed Method

Fig. 8 Pitch Contour(6dB)
(a)Reference Pitch (b) LPC Method
(c) Cepstrum Method (d) Proposed Method

5 Conclusion

The exact pitch extraction from speech signal is very difficult due to the effect of formant and transitional amplitude. So in this paper, the pitch is detected after the elimination of formant ingredients by flattening the spectrum in frequency region. The effect of the transition and change of phoneme is low in frequency region. Also, we proposed the new flattening method of log spectrum and the performance was compared with LPC method and Cepstrum method. This paper proposed the accurate pitch detection that the effect of formant can be removed by flattening the spectrum and the resolution of frequency also can be increased without increasing the number of FFT point.

Reference

[1] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech signals, Englewood Cliffs, Prentice-Hall, New Jersey, 1978.

[2] P. E. Paparnichalis, Practical Speech Processing Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1987.

[3] S. Seneff, "Real Time Harmonic Pitch Detection," IEEE Trans. Acoust. Speech, and Signal Processing, Vol. ASSP-26, pp. 358-365, Aug. 1978.

[4] S. D. Stearns & R.A. David, Signal Processing Algorithms, Prentice-Hall, Inc, Englewood Cliffs, New-Jersey, 1988.

[5] M. Bae, and S. Ann, "Fundamental Frequency Estimation of Noise Corrupted Speech Signals Using the Spectrum Comparison," J., Acoust., Soc., Korea, Vol. 8, No. 3, June 1989.

[6] M. Lee, C. Park, M. Bae, and S. Ann "The High Speed Pitch Extraction of Speech Signals Using the Area Comparison Method," KIEE, Korea, Vol. 22, No. 2, pp.13-17, March 1985.

[7] M. Bae, J. Rheem, and S. Ann "A Study on Energy Using G-peak from the Speech Production Model," KIEE, Korea, Vol. 24, No. 3, pp. 381-386, May 1987.

[8] Hans Werner Strube , "Determination of the instant of glottal closure from the speech wave," J., Acoust., Soc., Am, Vol. 5, No. 5, pp. 1625-1629, November 1974.

[9] M. Bae, I. Chung, and S. Ann, "The Extraction of Nasal Sound Using G-peak in Continued Speech," KIEE, Korea, Vol. 24, No. 2 pp. 274-279, March 1987.