A STUDY ON A PITCH ALTERATION BY USING THE FORMANT AND PHASE COMPENSATION TECHNIQUE

Won Park, Hyung-Bin Park, Myung-Jin Bae
Dept. of Telecom. Engr., Soongsil Univ.
Seoul 156-743, Korea
mjbae@saint.soongsil.ac.kr

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

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

10.21437/Eurospeech.1999-404

ABSTRACT

In the area of the speech synthesis techniques, the waveform coding methods maintain the intelligibility and naturalness of synthetic speech. In order to apply the waveform coding techniques to synthesis by rule, we must be able to alter the pitch of synthetic speech. In this paper, we propose a new pitch alteration method that compensates the formant and phase information by using the inverse filter of LP analysis and the pitch alteration method of the time domain. For performance test we used as the spectrum distortion rate as objective criterion and the MOS(Mean Opinion Score) was used as subjective criterion. As a result, the spectrum distortion and MOS are obtained by 0.6% and 4.0, respectively.

I. INTRODUCTION

Recently, owing to the rapid progress of VLSI technology, the 256 Mbit memory size per chip package is available in market. For the 32 kbps ADPCM waveform coding, such a long speech data as a half hour lasting speech can be stored by using one 64 Mbit chip. This makes the improvement of speech quality more important target than the reduction of memory size.

The waveform coding method or the hybrid coding method, also, is preferable to the speech synthesis techniques for high quality. Although, for a long time, the waveform coding method and the hybrid coding method have been used for sentence based synthesis in synthesis technique by analysis, they are not proper to syllable or phoneme based synthesis techniques, because of the difficulty in controlling the excitation source.

Therefore the waveform coding techniques for the synthesis by rule is relatively good method to maintain the naturalness and the intelligibility comparable to the original speech.

According to processing domain, the pitch alteration method is classified into three domains; time domain, frequency domain and time-frequency hybrid domain. There are multi-pulse method and pitch halving method in time domain. To alter the pitch period, Caspers and Atal proposed the method in which zeros are inserted or the data is deleted between pulses on MPLPC[3]. However, because the pulse train on MPLPC is related to pitch and formant, serious spectrum distortion occurs. Varga and Fallside had proposed the pitch extension method by LPC coefficients, which also causes serious spectrum distortion because they simply deleted a part of waveform when shortening the pitch[4].

In this paper, we propose a new pitch alteration method in which pitch-altered waveform can be obtained by combining the pitch data which come from the cepstrum analysis and the phase data which come from the time scaling pitch control method.

II. CEPSTRAL PITCH ALTERATION METHOD

Unlike in the source coding method, the pitch variation of the speaker must be known prior to change the pitch period in the waveform speech coding. This comes from the fact that the variations of the accent and the emotion of a speaker result in the variation of the pitch period around the average value of that. Especially, since the waveform coding method conserves the characteristics of a speaker and the message informations, its intelligibility is relatively good. So, it is needed to alter the pitch period according to the average pitch period which mainly appears in the speech signal of the speaker. Therefore, the precise pitch detection must be carried out prior to changing the pitch.

From the result of cepstral analysis for the voiced speech, the combined contributions of vocal tract, glottal pulse and radiation appear on the lower part of quefrency domain and decay rapidly for large quefrency. The remarkable peak corresponding to the excitation source appears around the pitch period on the higher quefrency domain. So, by inserting lifter around the pitch where the cepstrum decay to zero on the quefrency domain, we can separate the formant components and the fundamental informations. This is called as the cepstral analysis method[1].

Speech signal can be separated into magnitude component and phase component by Fourier transform. So, the magnitude component of the Fourier transformed speech signal is as follows:

$$S(k) = \int_{-\infty}^{\infty} s(n)\ e^{-j\frac{n}{2\pi N}k} dn \tag{1}$$

$$M(k) = 10 \log S^2(k) \tag{2}$$

To control the pitch in frequency domain, spectrum scaling is used. Spectrum must scale on the speech excitation spectrum. Thereby, the separation of component is performed before pitch alteration by the cepstral analysis.

If the formant components, $S^*(k)$, extracted by cepstral analysis are subtracted from M(k) as Equation (3), the flattened harmonics spectrum could be separated:

$$S_P(k) = M(k) - S^*(k) \tag{3}$$

Where $S_p(k)$ is the flattened harmonics spectrum. For this signal, the scaling rate in frequency domain is the inversion of the scaling coefficient of time axis.

$$\hat{S_P}(k) = S_P(k \times \rho^{-1}) \tag{4}$$
$$(k = 0, 1, 2, 3, \ldots, N-1)$$

In Equation (4), ρ^{-1} represents the frequency scaling rate, and $\hat{S_P}(k)$ expresses the changed harmonics spectrum. It must decrease the interval of the fundamental frequency by ρ^{-1} for expanding pitch, and increase by ρ^{-1} for compressing pitch.

Since the effect depending on the kind of window is serious, the beginning point of window has to be synchronized to the exciting point of the glottal pulse. For this, the phase information of the waveform must be kept unchanged while changing the pitch period, so, time domain pitch extraction is desirable.

In this paper, we adopt the area comparison method[2] in time domain. However, since the automatic pitch extraction is not positively necessary when editing the waveform for synthesis, semi-automatic pitch extraction and manual pitch extraction also may be a good adoption.

III. FORMANT AND PHASE COMPENSATION

In the cepstrum pitch alteration method proposed previously in [5], how phase information can be kept unchanged is the unresolved problem. So, we propose the phase compensation method in which we use the pitch alteration method of time domain. Also according to the inverse filter of LP analysis which is represented as a following Equation (5), $H^{-1}(z)$, we compensate the formant information.

$$H^{-1}(z) = \sum_{i=1}^{p} a_i\, z^{-i} \tag{5}$$

Prior to control the pitch in time domain, voiced speech signal is passed through the low pass filter(LPF) represented as a following Equation (6) with a cut-off bandwidth as a pitch period.

$$s'(n - \frac{N}{2}) = \sum_{i=0}^{N-1} s(n-1) \tag{6}$$

Where N is the cut-off bandwidth interval of LPF, because the cut off frequency, f_T, equals f_S/N. For the harmonics above the fundamental frequency is removed from the signal, the LPFed signals are similar to excitation source of the voiced signals. Now, the signal is scaled at time axis as follows:

$$\widehat{s(n)} = s'(n \times \rho) \qquad (7)$$

Where $\widehat{s(n)}$ is the scaled signal in time domain. s′(n) is the low pass filtered signal. The scaling factor is ρ as follows:

$$\rho = \frac{P'}{P} \qquad (8)$$

where P is a speaker's pitch and P′ is an expected pitch. If ρ is smaller than 1, we would obtain the signal with compressed pitch, Reversely if ρ is larger than 1, we would obtain the expended pitch. Then, the FFT is applied to the signal scaled at time axis.

As represented so far, the phase information is obtained from the FFT spectrum after we alter the pitch period of the speech in time domain by time scaling method, and then it is combined with the magnitude of the spectrum which is obtained by cepstrum pitch alteration method.

Ⅳ. EXPERIMENTAL RESULTS

The proposed algorithm has been implemented on the IBM PC(Pentium 233MHz) with the 16-bit AD-DA converter. The speech signal is low-pass filtered at 4 kHz and sampled at 11 kHz. Five phoneme balanced Korean sentences are used as test data. Each sentence is pronounced 5 times by three males and two female speakers. The following sentences are used in our experiment:

Data 1. /INSUNE KOMAGA CHUNJAE SONYUNWL JOAHANDA/
Data 2. /YESUNIMKESEO CHUNJICHANGJOWI KYOHUNWL MALSUMHASEOSSDA/
Data 3. /SOONGSILDAE JUNGBOTONG SHIN-KWA UMSENG SINHOCHURI YUNGUTEEMIDA/
Data 4. /KAMSAHAMNIDA/
Data 5. /May I Help You/

The analysis frame consists of 512 samples. First, the beginning point of pitch period is obtained by using the area comparison method to get the pitch interval which is needed for synthesis by rule in waveform coding. After repeating and zero padding the pitch interval to get a frame which consists of 512 samples, we altered pitch period of the residual signal which is gotten by inverse filter of LPC analysis. Figure 1 is the proposed block diagram in this paper.

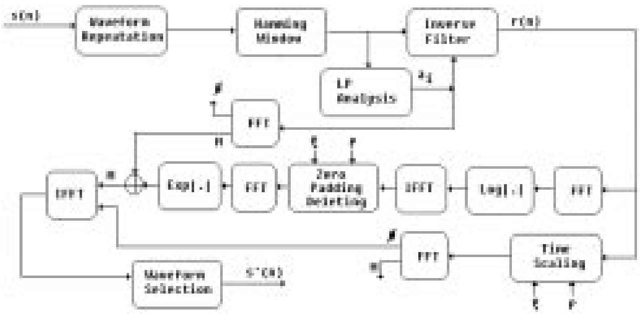


Figure 1. A block diagram of the proposed pitch alteration technique.

The examples of which depicted in Figure 2 are represented the results of Data 1 in pitch alteration ratio 150%. The result in Figure 2(c) is obtained by using the proposed method without losing the formant and phase information.

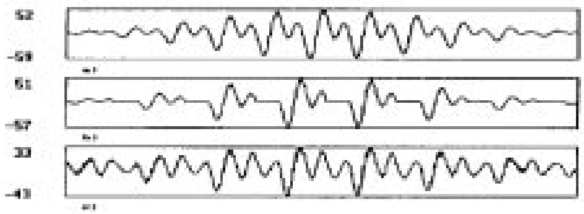


Figure 2. An example of pitch period extended by 150% by using the cepstrum pitch alteration method

Above the fact we can minimize the formant and phase distortion which is generated around the conjunction point of adjacent waveform in synthesis by rule using waveform coding.

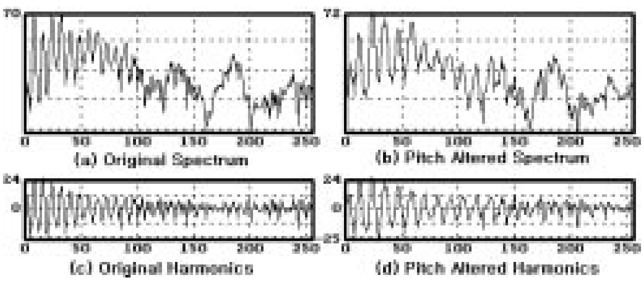
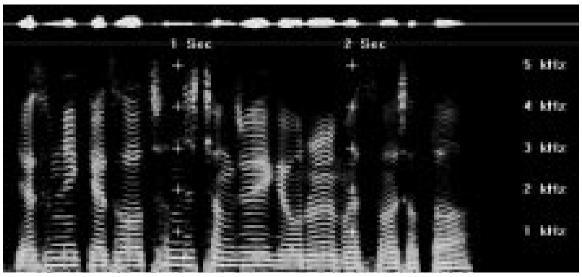
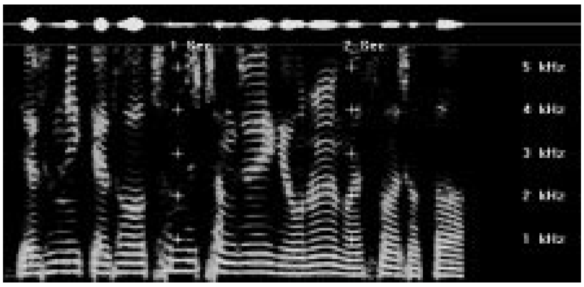


Figure 3. An example of pitch compressed by 70% by using the cepstrum pitch alteration method.



(a) Original speech signal and its spectrogram.



(b) Pitch altered speech signal and its spectrogram

Figure 4. The spectrograms of original speech signal and pitch altered speech signal by 70%.

Table 1. Spectrum distortion rates in the proposed method.

Alteration Rate	Female (%)	Male (%)	Average (%)
90% → 111%	0.21	0.17	0.19
80% → 125%	0.38	0.32	0.35
70% → 142%	0.66	0.48	0.57
60% → 166%	0.99	0.64	0.82
50% → 200%	1.39	0.74	1.07
Average	0.73	0.47	0.60

Table 2. The result of MOS test in the cepstral pitch alteration method.

Alteration Rate	MOS	
	Conventional method	Proposed method
90% → 111%	4.0	4.3
80% → 125%	3.8	4.2
70% → 142%	3.5	4.1
60% → 166%	3.1	3.9
50% → 200%	2.8	3.3
Average	3.4	4.0

V. CONCLUSIONS

Speech synthesis techniques are classified into three groups; waveform coding, source coding and hybrid coding. So far, in synthesis by rule the waveform and hybrid coding methods are mainly used to synthesis method by analysis because it is not separate the excitation and vocal tract parameter. However, if it is possible to alter the pitch period when the waveform coding is used, synthesis by rule is available for maintaining as intelligible and natural as the original speech. In this paper, we proposed the new pitch alteration method, in which the formant information compensate by using inverse filter of the LP analysis and the phase compensation is performed on that by using the time scaling method. When we alter the pitch period over the magnitude spectrum flattened on the frequency domain where the formant informations almost does not exist. Consequently we can minimize the magnitude spectrum distortion by using the inverse filter and the phase spectrum distortion which is generated in conjunction point of two analyzed frame when using synthesis by rule in waveform coding.

As a result, we can get the spectrum distortion of 0.6% and the MOS score of 4.0.

REFERENCES

[1] L.R. Rabiner & R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, New Jersey, 1978.

[2] M.J. BAE and S.G. ANN, "The High Speed Pitch Extraction of Speech Signals using the Area Comparison Method", KITE, Vol.2, No.2, pp.101-105, Feb., 1985.

[3] B.E. Caspers and B.S. Atal, "Changing Pitch and Duration in LPC Synthesised Speech using Multipulse Excitation", J. Acoust. Soc. Amer., Vol.73, No.1, pp.55, Spring, 1983.

[4] A. varga and F. Fallside, "A Technique for Using Multipulse Linear Predictive Speech Synthesis in Text-to-speech Type System", IEEE signal processing, Vol.ASSP-35, No.4, pp.586-587, APRIL 1987.

[5] MyungJin Bae, KyuHong Kim, Woncheol Lee "On a Cepstral Pitch Alteration Technique for Prosody Control in the Speech Synthesis System with High Quality", EUROSPEECH'97 Vol.2, p.609-612 Sep. 23, 1997.

[6] J.D. KIM, S.J. BAEK, M.J. BAE, "On a Pitch Alteration Technique in Excited Cepstral Spectrum for High Quality TTS," Proceedings of ICSLP'98, Vol., No., pp.-, Dec. 1998.