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

International Symposium on Chinese Spoken Language Processing (ISCSLP 2002)
Taipei, Taiwan
August 23-24, 2002

A NOVEL APPROACH FOR PITCH MODIFICATION ON TIME DOMAIN

LI Ming, WANG Zhiyu, WEN Yuan, HOU Zhen, YU Tiecheng

Speech Processing Lab, Institute of Acoustics, Chinese Academy of Sciences, Beijing
liming@mail.ioa.ac.cn

ABSTRACT

Pitch is an important element of speech property. Pitch modification is necessary for many speech processing applications such as speech synthesis. Methods of pitch modification can be categorized as methods on time domain and methods on frequency domain. In this paper we propose a novel approach which can modify pitch on time domain. This approach takes advantage of human auditory perception property that human is not sensitive with the low energy parts of speech. Utilizing this property we shorten or expand the pitch period by deleting or copying the low energy parts of speech. Our experiments show that this approach can change pitch period effectively.

1. INTRODDUCTION

As for the transformation between male speech to female speech, Chu Min presented a simple method on time domain in which she controlled the transformation by two parameters, LBOP(Low Boundary of Pitch Range) and SMF(Sample-rate Modification Factor)[1]. She utilized TD-PSLOLA[2] to modify pitch in her proposal. Her experiment results showed that she got good results for the transformation from female speech to male speech.

In this paper we present another pitch modification approach which is motivated by the speech production process. During the speech generation process, after every vocal cord vibration glottis waveform is filtered in vocal track and its envelope decreases gradually. In each pitch period the waveform part whose amplitude is lowest has little effect on speech perception. By taking advantage of this property, we can do some changes on this part of speech which can modify the pitch period without affecting auditory perception too much.

2. THE APPROACH OF PITCH MODIFICATION ON TIME DOMAIN

If we would like to change pitch period we have to know the pitch period first. We set up a pitch tracker module using Sub-Harmonics Summation method[3] which can output precise pitch values for most speech frames. Under the guide of pitch tracker it is easy to detect the beginning and end points of each pitch period.

In the following we will describe how to find that part of waveform that we want. First we should find the maximum peak in a pitch period. Then we detect the first zero across point from the peak in the time inverse direction. Finally we can locate a small part of waveform before this zero across point whose energy is lowest at the given length. This part of speech is just what we need. We can cut or stretch this part of speech to increase or decline the pitch frequency. The length of this part of waveform can be adjusted at the demand of pitch modification.

3. EXPERIMENT RESULTS

For male speech data we find lowest energy part in each pitch period. These lowest energy parts will be deleted and other speech waveform will remain unchanged. Figure 1 is an example that the pitch period of a male speech will be reduced. Each coterminous remained and cut region consist of an whole pitch period. The cut regions are the parts which is to be deleted while the remained regions are the ones which is to be unchanged. Figure 2 is the waveform whose pitch period has been reduced. For female speech data we insert some waveform during each pitch period, by which the pitch frequency is lowered. In Figure 3 the 'u' parts of waveform are the ones that do not change while the 'r' parts of waveform are the ones that will be appended by their copy ones. Figure 4 is the speech waveform which is derived from Figure 3 and whose pitch has been expanded.

Table 1 and Table 2 are the experiment evaluation results. P is the ratio of new pitch frequency to original pitch frequency. If P is greater than 1, pitch period is reduced. If P is smaller than 1 , pitch period is expanded. MOS(Mean Opinion Score) is the evaluation on the transformed speech. For male speech the lower P is, the transformed speech sounds more like a female speech and the lower the speech quality is. Likewise for female speech the higher P is, the transformed speech is more like a male speech.

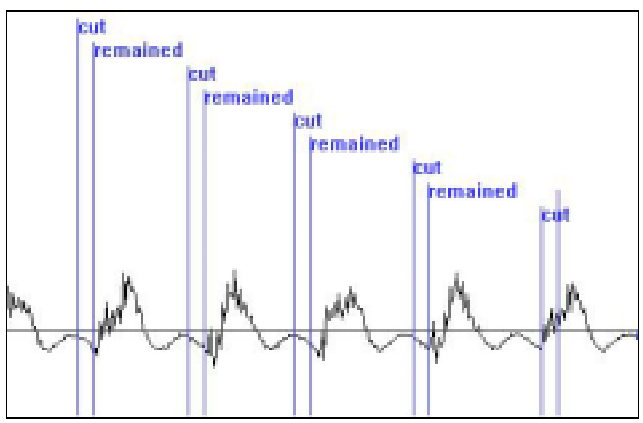

Figure 1. Part of male speech whose pitch period is to be shortened

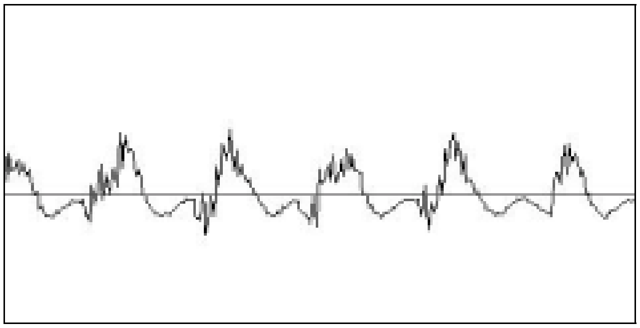

Figure 2.Part of male speech whose pitch period has been shortened

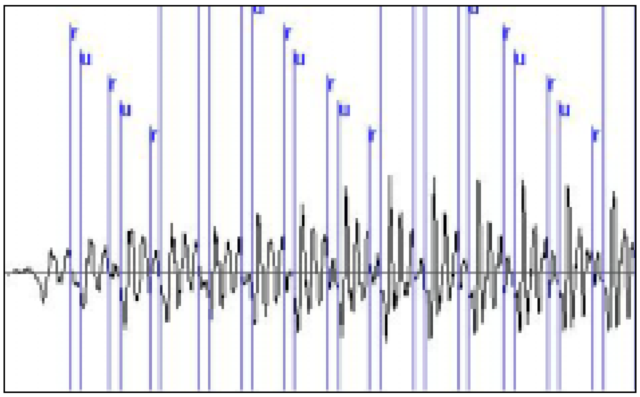

Figure 3. Part of female speech whose pitch period is to be expanded

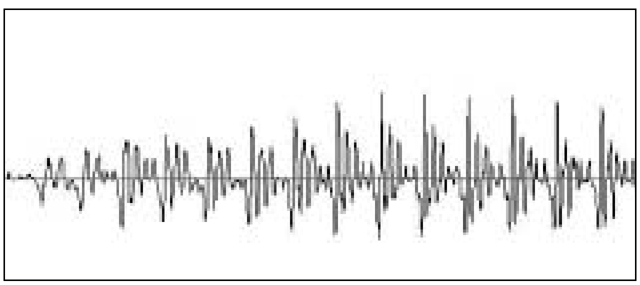

Figure 2.Part of female speech whose pitch period has been expanded

Table 1. *Lowering pitch frequency for female speech*

P	0.6	0.7	0.8	0.9
MOS	3.8	3.9	4.0	4.0

Table2. *Raising pitch frequency for male speech*

P	1.3	1.4	1.5	1.6
MOS	3.9	3.8	3.8	3.8

4. SUMMARY

This paper presents a novel approach for modification of pitch on time domain. It explores the human auditory property to reduce or expand pitch period. This method is quite simple and easy to implement. Its effectiveness is verified by our experiment results.

5. REFERENCES

[1] Chu Min, Lv Shinan, "A High quality Text-To-Speech System for Chinese", Journal of Acoustics(Supplement) 1995, Vol. 21

[2] Charpentier.F., Moulines.E. (1990), "Pitch-synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones", Speech Communication, 9, pp. 453-467

[3] Dik Hermes, "Measurement of pitch by subharmonics summation", *Journal of Acoustics of Society of America*, Am 83(1), Jan.. 1988, pp. 257-264