

# On a Pitch Duration Technique for Prosody Control

JongKuk Kim, HernSoo Hahn, Uei-Joong Yoo, and MyungJin Bae

**Abstract**—In this paper, we propose a method of alter duration in frequency domain that control prosody in real time after pitch alteration. If there has a method to alteration duration freely among prosody information, that may used in several fields such as speech impediment person's pronunciation proof reading or language study. The pitch alteration method used control prosody altered by PSOLA synthesis method which is in time domain processing method. However, the duration of pitch alteration speech is changed by the frequency domain. In this paper, we altered the duration with the method of duration alteration by Fast Fourier Transformation in frequency domain. Consequently, the intelligibility of the pitch and duration are controlled has a slight decrease than the case when only pitch is changed, but the proposed algorithm obtained the higher MOS score about naturalness.

**Keywords**—PSOLA, Pitch Alteration, Duration Control.

## I. INTRODUCTION

**I**N general, prosody information such as pitch, energy, and duration so on., is in speech. So if prosody information is applied to synthesized speech when the sentence is synthesized, more accurate information can be communicated. In this paper, pitch alteration processed as the method of controlling the prosody in time domain and the duration altered in frequency domain. First of all, in case of altering the pitch, pitch alteration by PSOLA synthesis method applied [8].

In order to maintain the characteristic and intelligibility of speaker pitch alteration is based on the pitch obtained from speakers. Therefore, the initial point of speaker should be detected to implement the pitch alteration. Also, at speech analysis detecting correctly pitch point pitch cycle emphasize that synthesized speech because can analyze that is done, and can grasp phase characteristic of excitation source at synthesis get can. Detection of initial pitch point used by linear prediction analysis used in the paper and pitch is altered by synthesis method using by the information of initial pitch

JongKuk, Kim is with the Department of Information & Telecommunication at Soongsil University, Seoul, 156-743, Korea. He is currently the researcher professor (corresponding author to provide phone: 82-2-824-0906; fax: 82-2-824-0906; e-mail: kokjk91@ssu.ac.kr).

HernSoo, Hahn is with the Department of Information & Telecommunication at Soongsil University, Seoul, 156-743, Korea. He is currently professor (e-mail: hahn@ssu.ac.kr).

Uei-Joong, Yoo is with with the Department of Biomedical Engineering, Gachon University of Medicine and Science, Incheon, 406-799, Korea. He is currently a professor (e-mail: ujyoon@gachon.ac.kr).

MyungJin, Bae is with the Department of Information & Telecommunication at Soongsil University, Seoul, 156-743, Korea. He is currently professor (e-mail: mjbbae@ssu.ac.kr).

point[7]. Duration by alter of frequency domain of pitch altered speech by PSOLA synthesis method of pitch cycle unit is altered over. So changed duration in the speech should be needed to synchronize with the original duration of speech.

In this paper, we altered the duration with the alteration method of duration using by Fast Fourier Transformation in frequency domain. If can alter duration freely among prosody information, may use in several fields such as speech impediment person's pronunciation proofreading or language study studying.

## II. PITCH ALTERATION TECHNIQUE

When convert prosody information, pitch detection process should be achieved beforehand. However, average pitch information between analysis frame is difficult to reflect well phoneme change characteristic and so on in speech signal. Therefore, can analyze pitch cycle if can detect correct pitch point and transformation of prosody information is easy. This paper used pitch point detection method to need in prosody control that proposes in laboratory which see.

First, apply amplitude characteristic of glottal and cycle characteristic that is neted by analysis section after pass adversely to filter that was expressed by linear prediction coefficient that number area of high frequency through pre-emphasis filter is emphasized and achieved detection process pitch point of time [6]. Then, altered pitch by PSOLA synthesis method using such got pitch point information. Finally, complete content and organizational editing before

## III. PITCH DURATION CONTROL

This paper wishes to propose about method that change duration by real time without alter of voice color using Fast Fourier Transformation. This method altered duration without alter of spectrum by reduce calculation time using FFT by duration alteration method in frequency domain and achieve IFFT process by FFT point's 2<sup>n</sup> point after achieve 2<sup>n</sup> 's Interpolation and Decimation each amplitude and phase.

The next Fig. 1 displays block diagram of duration alteration method that use Fast Fourier Transformation. Then, achieve cycle contraction Interpolation and Decimation process by 2<sup>n</sup> such neted each amplitude and phase component. Then, had achieved IFFT by 2<sup>n</sup> point and get speech that duration is altered. The next Fig. 2 appears about the example that alter duration using Fast Fourier Transformation that propose in paper that see in one frame, and Fig. 3 displays waveform that alter sentence whole duration by picture.

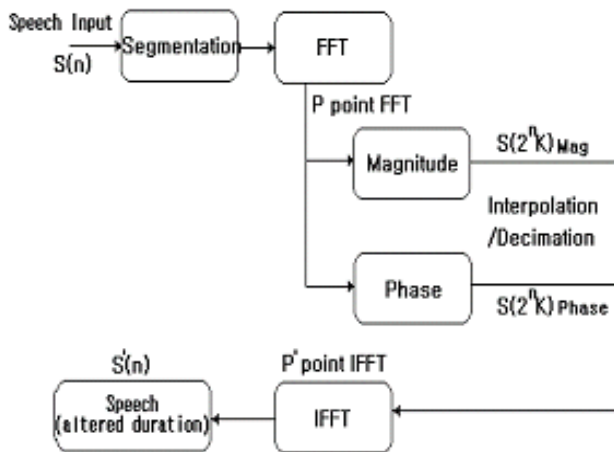


Fig. 1 Duration control Block diagram that use FFT.

Fig. 4 displays each Pitch Frequency Contour. As appeared in figure, because Pitch of when altered duration sees that is kept changelessly by average 150Hz, voice color can know that do not alter and alters length.

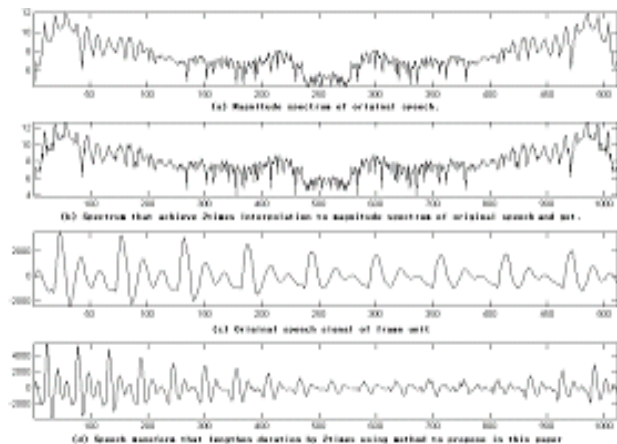


Fig. 2 Example that change duration using FFT in one frame

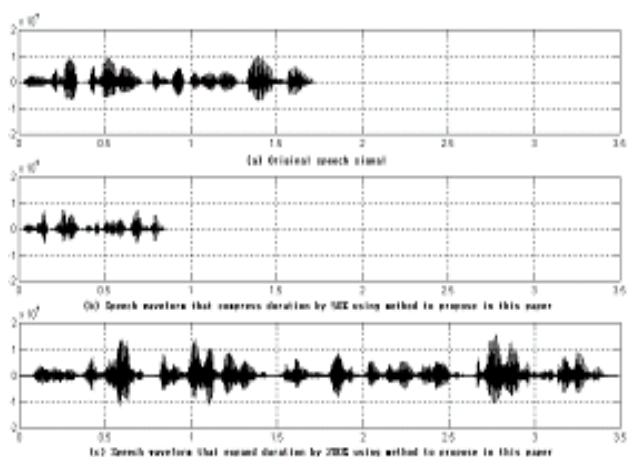


Fig. 3 Example that change duration using proposed method

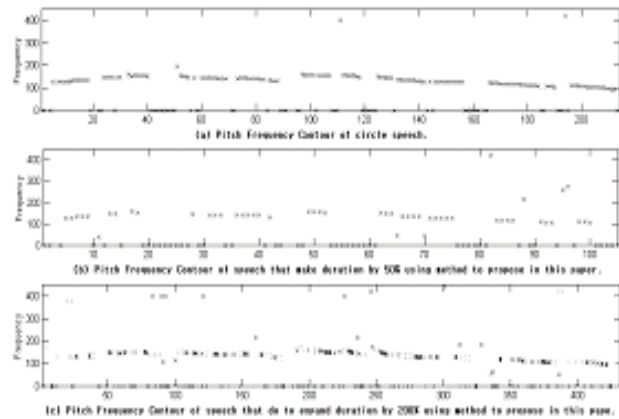


Fig. 4 Pitch contour that change duration using proposed

#### IV. EXPERIMENTAL RESULTS

Method to propose this paper that see an experiment and result to do simulation to IBM-PC/586 microphone input interfacing possible 16 bit A/D converter next time speech utterance through 2 man and 2 woman speakers to make voice do and stored this quantizing 16 bit by 11kHz's the sampling rate. Because do length of frame by 45ms about each utterance in pitch alteration, achieved frame overlap process as estimate degree (p) and duration alter that propose did length of frame to 256 Samples. Used next representative sentences by utterance to search incidental and performance of value.

- 1: /Insune komaneun cheonjaesonyuneul joahanda/
- 2: /Soongsildae  
jeongbotongshin-gonghakkwa eumseiongtongshin  
yeungutimida/
- 3: /Yesunimkeoseocheonjichangjoeu  
kyohuneul malseumhasyuda./
- 4: /Gongilisamsaoyukchilpalgu/

Embodiment achieved in C-language and MATLAB to embody method to propose in this paper. Simulation applied pitch alteration method by PSOLA composition way according to the alter rate after get pitch point ten using detecting means automatic pitch point of time that propose in laboratory which see[8]. Also, duration alter used duration alteration method in frequency domain that use Fast Fourier Transformation that propose in this paper.

The next Fig. 5 displays block diagram of real time prosody control system. Block diagram Fig. 6, 7 of real time prosody alter system, shows example of waveform that alter speech that become pitch alteration using above method by duration that original speaker makes speech. Block diagram Fig. 6, 7 of real time prosody alter system, shows example of waveform that alter speech that become pitch alteration using above method by duration that original speaker makes speech.

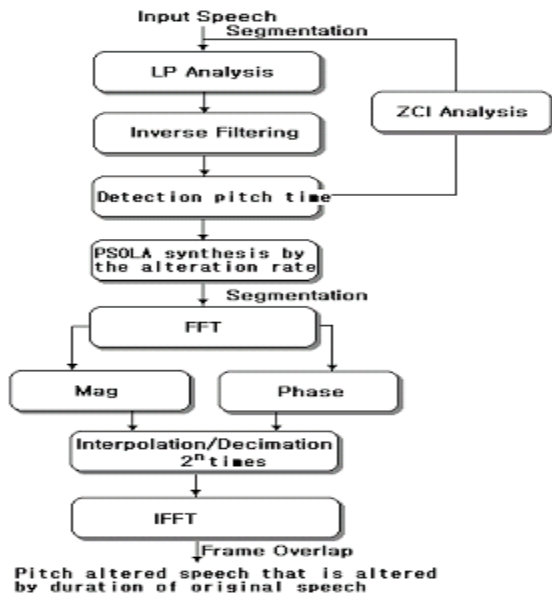
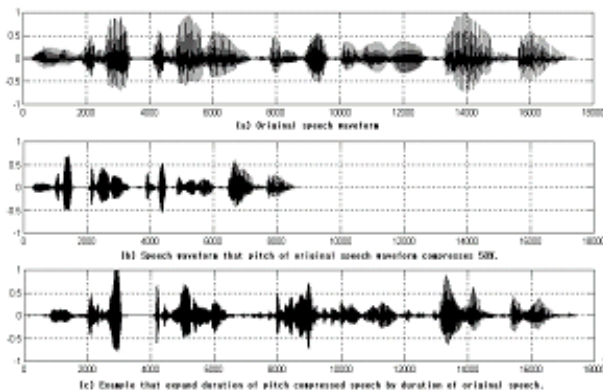
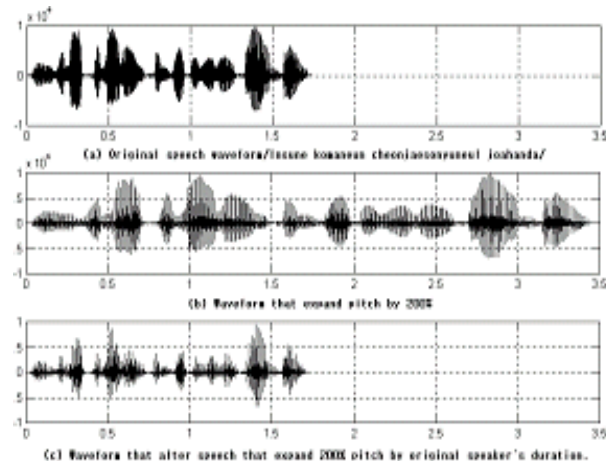


Fig. 5 Block diagram of real time duration control

Fig. 6 After compress pitch by 50%  
/ Insune komaneun cheonjae sonyuneul joahanda/Fig. 7 After refurbish pitch by 200%  
/Insune komaneun cheonjae sonyuneul joahanda/

This paper used pitch alteration method by automatic pitch point detection that research reactor pitch alteration about real time prosody control proposes in laboratory, and duration alteration method used duration alteration method in frequency domain that use Fast Fourier Transformation that propose in paper.

An experiment used proper move times ( $2^n$ ) alter when change duration because use Fast Fourier Transformation. Therefore, experimented because do 200% extension when extend duration and used 50% zip when compress. Paper achieved subjective sound quality estimation for naturalness Performance estimation because use speech that change pitch and Fast Fourier Transformation that propose in paper, come intelligibility with speech that change duration.

The next Table I tells to 5 listeners who is choosed random and display definition and naturalness measuring average MOS (Mean Opinion Score) about speech utterance. MOS's grade divided by Excellent (5), Fair (4), Good (3), Poor (2), Unsatisfactory (1)

TABLE I  
MOS TEST FOR PROSODY IN SPEECH

Listeners	Prosodic result[MOS]			
	Only pitch alteration		Both pitch and duration alteration	
	Intelligibility	Natural	Intelligibility	Natural
Listener1	3.70	3.20	3.55	3.60
Listener2	3.65	3.10	3.45	3.50
Listener3	3.60	3.25	3.40	3.40
Listener4	3.80	3.30	3.40	3.35
Listener5	3.85	3.40	3.60	3.50
Average	3.72	3.25	3.48	3.47

When compared naturalness with intelligibility of speech that change duration by pitch altered speech and duration alteration method that propose with result that appear Table 1, speech that alteration pitch liked fairly definition but naturalness came out as is bad relatively because altered with duration by change of frequency domain when alteration pitch. However, speech that alteration pitch changed speech by original speaker's duration could get superior synthesized speech relatively than speech that alteration pitch when compare with when altered pitch and evaluated naturalness in extent that do not injure greatly definition of sound quality.

## V. CONCLUSION

Pitch alteration changed using PSOLA synthesis method that is time domain processing law by one of method to regulate prosody in this paper [8]. If keep speaker's character and intelligibility in time domain pitch alteration, important pitch of speaker by standard pitch alter consist must. Therefore, must be able to detect the speaker's pitch point to achieve pitch alter. And altered pitch by PSOLA synthesis method because using neted so pitch point information. However, duration by change of frequency domain of pitch alteration speech is altered. So changed duration in the speech should be needed to

synchronize with the original duration of speech. altered duration by duration alteration method in frequency domain that use Fast Fourier Transformation in this paper therefore. Experiment result of this paper original speech with intelligibility of case that alter pitch and speech that alter duration naturalness compare.

Intelligibility of when altered pitch in an upside experiment could get MOS Score of average 3.68, and naturalness could get MOS Score of average 3.18. Also, pitch intelligibility of case that alter altered speech by original speaker's duration could get MOS Score of average 3.40 and naturalness could get average 3.38 MOS Scores.

As a result, came out as is bad than case that intelligibility of speech that alter duration alters pitch, but naturalness side could get better result. Is going to need research about duration that is not times through DFT by hereafter.

#### REFERENCES

- [1] G. Bristow, Electronic Speech Synthesis, McGraw-Hill, 1984.
- [2] E.J. Yannakoudakis, P.J. Hutton, Speech Synthesis and Recognition Systems, Ellis Horwood Ltd., 1987.
- [3] J.R. Deller, J.G. Proakis, J.H.L. Hansen, Discrete-Time Processing of Speech Signals, Macmillan Publishing Co., 1993.
- [4] H.G. Lee, M.J. Bae, U.C. Im, "Detection pitch point of speech signal by G-Peak detection", No. 6 signal processing combination art and science contest, the sixth book, 1, pp.58-61, 1993.
- [5] H.B. Park, W.R. Jo, J.D. Kim, W. Park, D.S. Shim, M.J. Bae, "Research about pitch alteration method of optimistical by the pitch alteration rate", No. 15 speech communication and new good handle workshop collection of learned papers, Vol.15, No.1, PP.460-464, August, 1998 21 - 22 day.
- [6] H.B. Park, M.J. Bae, "Research about pitch point detection for voice color alter", South Korea sound learned society, summer art and science announcement contest, emperor 19 book 1 (s) lake, No.1, pp.149-152, July, 2000 7 - 8 day.
- [7] B.E. Caspers, B.S. Atal, "Changing Pitch and Duration in LPC Synthesised Speech using Multipulse Excitation," J. Acoust. Soc. Amer., Vol.73, No.1, pp.55, 1983.
- [8] T. Takagi, E. Miyasaka, "A Speech Prosody Conversion System with a high Quality Speech Analysis-Synthesis Method," Proc. EUROSPEECH'93, pp.995-998, September 1993.

**JongKuk Kim** received the Ph.D. degree in Electrical Engineering from Soongsil University in 2005. He is currently the researcher professor at Soongsil University. His research interests include speech synthesis, speech recognition, speech coding, and audio coding.

**HernSoo Hahn** received the Ph.D. degree in Computer Engineering from University of Southern California in 1991. Currently, he is a professor. He has published papers on sensor development and applications on various fields. His research interests include application of ultrasonic and vision sensors to mobile robots and measurement systems.

**Uei-Joong Yoon** received the B.S. and Ph.D. degrees from Soongsil University, Seoul, Korea, in 1988 and 1993, respectively, all in the Department of Electronic Engineering. Since 2008, he has been with the Department of Biomedical Engineering, Gachon University of Medicine and Science, Incheon, Korea, where he is currently a Professor. His research interests are in the fields of the EMI/EMC and Digital Signal Processing.

**MyungJin Bae** received the Ph.D. degree in Electrical Engineering from Seoul National University in 1987. He is currently the Professor of the Dept. of Information & Telecommunication at Soongsil University. His research interests include speech synthesis, speech recognition, speech coding, and audio coding.