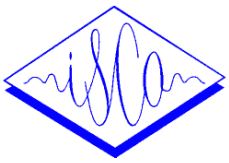


3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

# EFFECTS OF THE PHASE JITTERS ON NATURALNESS OF SYNTHESIZED SPEECH

Yun-Keun Lee and Seung-Kwon Ahn

Advanced Technology Lab.4, GlodStar Central Research Laboratory
Seoul, Korea

## ABSTRACT

Speech synthesizers based on frequency domain analysis usually have problems when they don't use phase information. For instance, they generate monotonous and machine-like speech. It has been found that phase jitters(PJs) are important factors on naturalness of synthesized speech. We analyze the PJs of natural speech using pitch synchronous FFT and construct the PJ model from this analysis. We also demonstrated that the synthetic speech using power spectrum envelope(PSE) and the PJ components can be almost indistinguishable from the natural speech.

*Keywords:* phase jitters, power spectrum envelope, pitch synchronous FFT

10.21437/Eurospeech.1993-215

## 1. INTRODUCTION

The speech synthesis methods that are widely used in frequency domain are LPC, LSP, PSE, Formant. These methods have focused in the modeling of spectral envelope and have not considered phase information [1][2]. So, many of these synthesis methods generate some unnatural monotonous speech, in particular for female speech.

In this study, we observe the effects of PJs on synthesized speech using PSE synthesizer. To do this, following experiments were performed. First, natural female speech signals of Korean vowels were recorded and PSEs were obtained by frame synchronous cepstrum analysis. Second, phase components were obtained by pitch synchronous fast Fourier transform(PSFFT) and mean and standard deviation of the phase were calculated. Third, the PJs were modeled and implemented using random signal generator and applied to PSE synthesizer. Fourth, we compared the quality of following three kinds of speech signals; (1) natural speech (2) synthesized speech by zero phase reconstruction (3) synthesized speech by jittered phase reconstruction. We have compared the spectrograms and performed opinion tests. As a result, it has been shown that the PJs are important factors on naturalness of synthesized speech.

## 2. THE PSE SYNTHESIZER

The speech synthesizer using PSE is not so popular as LPC or formant synthesizer. But to evaluate the effects of the PJs on synthetic speech, we use the PSE synthesizer that guarantees good quality. The basic block diagram of the PSE synthesizer is shown in Fig.1.

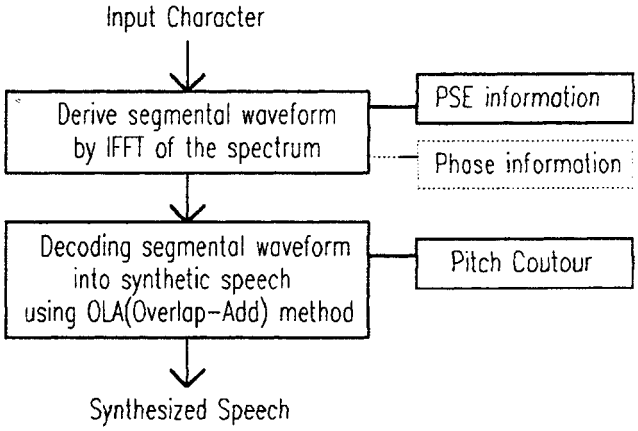



Fig.1 The block diagram of the synthesizer

As shown in Fig.1, the phase information(dotted block) is not considered and only the magnitude of the spectrum is used to synthesize the speech. So, we can get symmetrical speech waveforms by IFFT of the PSE and they are decoded into the synthetic speech using the overlap-add method as shown in Fig.2 [4][5].

Fig.6 shows spectrograms of the synthesized speeches and the natural speech. There are differences between (a) and (c). For example, the dithering effect in the spectrogram of the natural speech is not shown in the synthesized speech. The reason for this is that the PJs are eliminated by the zero phase reconstruction of the speech. This makes synthetic speech buzzy, monotonous and

unnatural [6]. In the next section, we analyze the PJs of the natural speech.

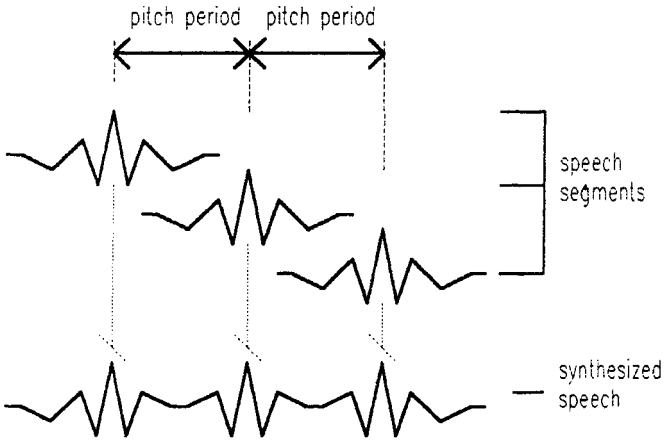



Fig.2 Decoding symmetrical speech waveforms into synthetic speech

## 3. ANALYSIS OF PHASE JITTERS

The pitch synchronous FFT can derive the phase information of each pitch period. We analyze the PJs of seven Korean vowels pronounced by a female speaker. The mean and the standard deviation(SD) of the phases of each frequency component are calculated as shown in Fig.3.,

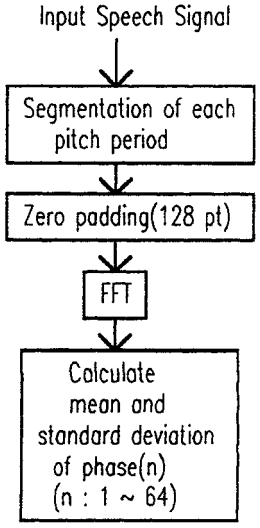



Fig.3. Analysis of PJs

The PJs are generated by not only the speech itself but the sampling of the signal that quantizes the phase. The phase we analyze above includes the PJs generated by the sampling. The pure PJs of the speech can be analyzed by the equations below. If we assume that the PJs generated by the sampling are uniformly distributed random signal, the variations of the PJs are given by

$$\sigma_s^2(f) = 1/(2M(f)) \int_{-M(f)}^{M(f)} \theta^2 d\theta = (1/3) \times M(f)^2 \quad (1)$$

where $\sigma_s(f)$ is the SD of PJs generated by the sampling and $M(f)$ is maximum phase deviation of $f$ Hz frequency component. $M(f)$ is given by

$$M(f) = (\pi \times f) / f_s \quad (2)$$

where $f_s$ is the sampling frequency. From (1) and (2), we have

$$\sigma_s^2(f) = (\pi^2/3) \times (f/f_s)^2 . \quad (3)$$

If we assume that the PJs of the speech and the PJs generated by the sampling are uncorrelated, the variations of the pure PJs of the speech are given by

$$\sigma_v^2(f) = \sigma_t^2(f) - \sigma_s^2(f) = \sigma_t^2(f) - (\pi^2/3) \times (f/f_s)^2 \quad (4)$$

where $\sigma_t(f)$ is the SD of the total PJs, $\sigma_v(f)$ is the SD of the PJs of the speech, and $\sigma_s(f)$ is the SD of the PJs generated by the sampling. From the analysis results and (4), we get the SD of the PJs of the pure speech. Fig.4 shows that the SD of jitters increases as frequency increases.

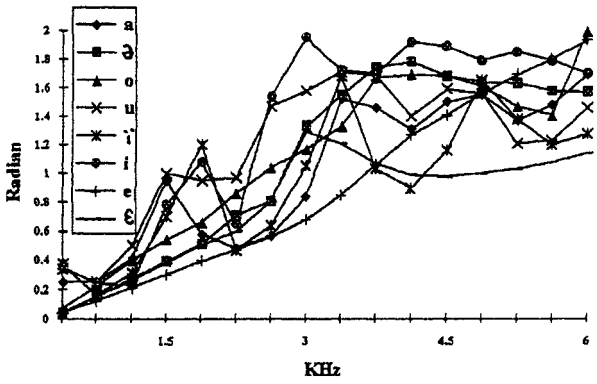



Fig.4 Standard deviation of jitters of seven Korean vowels

## 4. MODELING OF PHASE JITTERS

It is desired to implement a PJ generator to improve the quality of synthetic speech. We implemented the PJ

generator using random signal generator. The maximum amplitude of the random signal is equal to maximum phase deviation which can be calculated by equations below. From (1) the maximum phase deviation of the PJs are given by

$$M(f) = \sqrt{3} \times \sigma_v(f) \ . \tag{5}$$

The standard deviation of PJs versus frequency in Fig.4 can be approximated by a line equation

$$\sigma_v(f) = slp \times f \tag{6}$$

where $slp$ is the slope of the line. We calculated the slope by least mean square estimation(LMSE) and obtained 0.34 rad/KHz. From (5) and (6), we have

$$M(f) = \sqrt{3} \times slp \times f \ . \tag{7}$$

In the next section, we observe the quality of the synthesized speech including the phase jittering effect generated by the presented method.

## 5. EXPERIMENTAL RESULT

The synthesizer used in this experiment was implemented by computer simulation using IBM compatible PC. The block diagram of the synthesizer is presented in Fig.1. We put the PJ generator modeled in section 4 into the dotted block in Fig.1, which generates zero mean and jittered phase information. We recorded the speeches of 7 Korean vowels pronounced by a female speaker and A/D converted with 12kHz sampling rate and 12bits quantization. Each speech was segmented into frames and PSEs were extracted by the cepstral method as shown in Fig.5

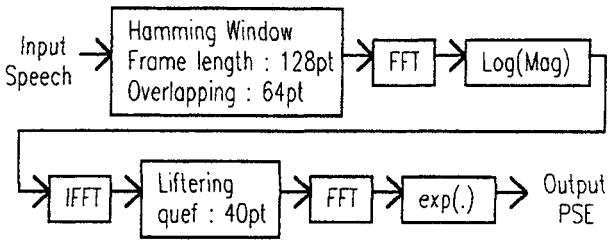
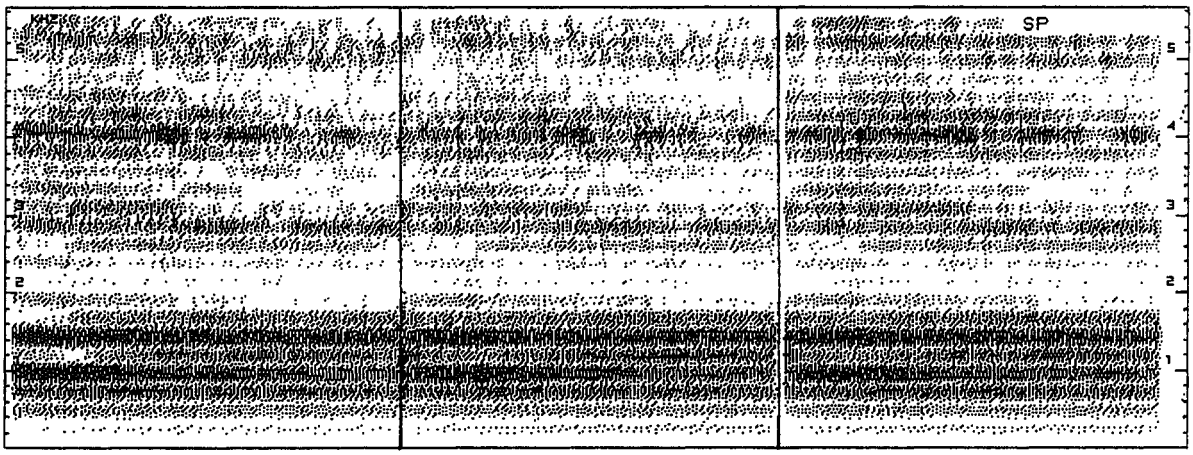



Fig.5 Flow chart of the derivation of PSE

Using the synthesizer and the PSEs of the vowels presented above, we synthesized two kinds of vowels by zero phase reconstruction and jittered phase reconstruction. The spectrograms of the synthesized vowels by the two methods and the spectrogram of the natural vowel are compared in Fig.6. This suggests that PJs are important information to improve the quality of the synthetic speech.

The opinion tests were conducted to assess the quality of the synthetic speech. The seven Korean vowels, that were generated by the two methods, were recorded. Twenty listeners listened to the recorded vowels presented in randomized order and scored the relative quality of the synthetic speech to the original speech in terms of naturalness. Naturalness was graded to 5 levels: 5(very good), 4(good), 3(fair), 2(bad), 1(very bad). The mean opinion score(MOS) is shown in Fig.7.




(a) Natural speech (b) Jittered phase reconstruction (c) Zero phase reconstruction

Fig.6 Spectrograms of Korean vowel /a/

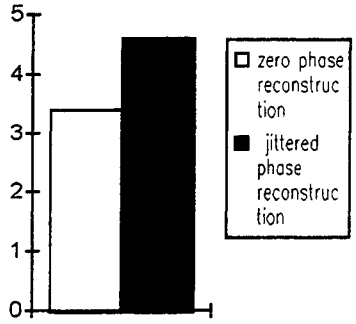



Fig.7 Results of the opinion tests(MOS).

## 6. CONCLUSION

The effects of the PJs on the quality of the synthetic speech were observed. The PJs of seven Korean vowels were analyzed and the standard deviations were calculated. With some mathematical assumptions the PJs were modeled and the PJ generator was implemented with the random signal generator and applied to the PSE synthesizer. Seven Korean vowels were synthesized by the PSE synthesizer in two ways - zero phase reconstruction and jittered phase reconstruction - and the spectrograms were compared with those of the natural speeches. This shows that the dithering effects, that are shown in the spectrograms of the natural speeches, are well reconstructed by the PJs. The opinion tests ware performed and the results showed that the quality of the synthetic speech can be improved by the phase jittering.

## REFERENCES

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

[2] S.Furui, Digital Speech Processing, Synthesis, and Recognition. Marcel Dekker,Inc. 1989

[3] T.Nakajima and T.Suzuki, "Power spectrum envelope(PSE) analysis-synthesis system", J.Acoust.Soc.Jpn., Vol.44, No.11, pp,824-832, Nov. 1987

[4] T.Yazu and K.Yamada, "The speech synthesis system for an unlimited Japanese vocabulary", ICASSP'86, Tokyo, Vol.2, pp.2019-2022, APR. 1986

[5] F.Charpentier and M.Stella, "Diphone synthesis using overlap-add technique for speech waveform concatenation", ICASSP'86, Tokyo, Vol.2, pp.2015-2018, APR. 1986

[6] F.Charpentier and E.Moulines, "Text-to-speech algorithms based on FFT synthesis", ICASSP'88, NewYork, Vol.1, pp.667-670, APR. 1988

[7] T.Moriya and M.Honda, "Speech coder using phase equalization and vector quantization", ICASSP'86, Tokyo, Vol.2, pp.1701-1704, APR. 1986