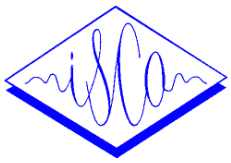


First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

A SPEECH ANALYSIS METHOD BASED ON A GLOTTAL SOURCE MODEL

Keiichi FUNAKI and Yukio MITOME

C&C Information Technology Research Laboratories, NEC Corporation,
4-1-1 Miyazaki, Miyamae-ku, Kawasaki, 213 JAPAN

ABSTRACT

The speech analysis method based on a glottal source model is important for efficiently constructing a rule-based speech synthesis system, since it is easy to control parameters, it can obtain smooth spectral trajectories and it can reduce memory capacity to store the parameters. This paper proposes a new speech analysis method based on the glottal source model. The proposed method uses glottal source and vocal tract inverse filterings in the frequency domain, and the model parameters are calculated so as to minimize inverse filtered error. The method can significantly reduce the computation amount and can assure vocal tract filter stability. Speech analysis synthesis experiments were carried out. The experimental results show that the speech spectrum is estimated more accurately and the estimated pole trajectory is smoother in the proposed method than in the conventional one. Further, synthetic speech is natural and intelligible. The results indicate that the proposed method is suitable for a rule-based speech synthesis system.

1 INTRODUCTION

Recently, a speech waveform concatenation method[1] and a residual wave controlled method[2] have been developed in a rule-based speech synthesis. These methods can synthesize intelligible speech. However, these methods sometimes produce noisy sound, such as clicks on speech unit boundaries due to the spectral discontinuity. Further, these methods require much memory to store the speech units, since the speech waveform or the residual signal as well as synthetic filter parameters are stored in the memory for every speech unit.

It is suitable for rule-based speech synthesis that the speech parameter concatenation method is adopted for a speech production model based on a glottal source model and a vocal tract filter model. The reason is as follows. This speech production model can approximate speech spectra much precisely than the conventional LPC does, and the waveform shape and its spectrum shape of a glottal source model can be controlled easily and flexibly. The vocal tract filter parameters can be interpolated in formant domain, so that noisy sound will not be generated. Moreover, glottal source parameters and vocal tract parameters require less amount of memory than speech waveform data or residual wave and LPC parameter data.

In this method, an automatic analysis technique, which extracts glottal and vocal tract parameters from human speech, is important for efficient development of high quality speech synthesis system.

Several analysis methods based on a glottal source model, which can estimate glottal and vocal tract parameters simultaneously and automatically, have already been proposed [5],[6].

However, these analysis methods involve the following two problems: A large amount of computation is required to calculate the model parameters, and there is no assurance for vocal tract filter stability.

In order to solve these problems in the conventional methods, the authors propose a new speech analysis method based on the glottal source model. In the proposed method, the glottal source characteristics can be removed by glottal source inverse filtering in the frequency domain, which differs from the so-called glottal inverse filtering[7]. The parameters of the glottal source and vocal tract models are calculated so as to minimize the inverse filtered error.

This paper is organized as follows: In Section 2, the conventional glottal LPC method[6] is described. In Section 3, the proposed analysis method is described. In Section 4, the experimental results using natural speech uttered by an adult male are presented, and the advantages of the proposed method are confirmed.

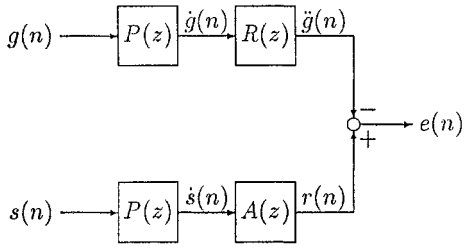
2 CONVENTIONAL GLOTTAL LPC METHOD

Several analysis methods, based on a glottal source model, have already been proposed such as GARMA method[5] and glottal LPC method [6]. The analysis model for glottal LPC method is shown in Figure 1 and minimization criteria(E) to determine the glottal and vocal tract parameters in the method is as follows.

$$E = \sum_n e(n)^2 = \sum_n [\sum_{i=0}^{p} \alpha_i \dot{s}(n-i) - \alpha_{p+1} \ddot{g}(n)]^2 \ , \alpha_0 = 1 \quad (1)$$

where $\alpha_i, p, \alpha_{p+1}, \dot{s}(n)$ and $\ddot{g}(n)$ are i-th AR parameter, the order of all-pole filter, the gain factor of the glottal wave $g(n)$, a pre-emphasized speech signal and a pre-emphasized signal for the glottal wave included the lip radiation factor, respectively.

In glottal LPC and GARMA method, it is necessary to pre-emphasize both the speech signal and the glottal wave in order to achieve pre-whitening of the inputs for the lip radiation and the inverse all-pole filter, and vocal tract and glottal parameters are identified by matching two signals, $r(n)$ and $\ddot{g}(n)$. Glottal and vocal tract parameters are obtained by the minimization of the error power ($\sum e(n)^2$) in the time domain. Since this minimization is sensitive to position of the glottal wave [5], these methods require accurate positioning for the glottal wave. Thus, it is necessary to solve several equations in order to determine the position of the glottal wave, and therefore, these methods require large amount of computation. In addition, since vocal tract parameters(AR and MA coefficients) are estimated by solving the covariance equation, filter stability are not always assured. Thus, the conventional methods involve two problems, (1)a large amount of computation is required, and (2)there is no stability assurance for the vocal tract filter.



$e(n)$: Analysis error
$R(z)$: Lip radiation filter
$P(z)$: Pre-emphasis filter
$A(z)$: Inverse all-pole filter

Fig.1 Speech analysis model for glottal-LPC method.

3 PROPOSED ANALYSIS METHOD

3.1 An Analysis Principle

Figure 2 shows an analysis model used in the proposed method[11]. The analysis model consists of two cascaded inverse systems, which are a glottal source inverse system($1/G(z)R(z)$) and a vocal tract inverse system($1/V(z)$). The inverse glottal source system is implemented by an inverse filter (a glottal source inverse filter), which has the estimated inverse characteristics of the glottal source. If the inverse characteristics for a glottal source is accurately estimated, glottal source frequency characteristics is removed from the speech signal by the filter, and the inverse filtered signal($u(n)$) can be effectively represented by a vocal tract model such as AR model. A vocal tract filter can be accurately estimated by analyzing the output signal($u(n)$) using conventional speech analysis method. In the proposed method, minimization criterion(E) to determine the glottal and vocal tract parameters is as follows.

$$E = \frac{1}{2\pi j} \oint_{|z|=1} |\frac{S(z)}{G(z)R(z)V(z)}|^2 \frac{dz}{z} \qquad (2)$$

where $S(z)$, $G(z)$ and $V(z)$ are $z-$transform of $s(n)$, $z-$transform of $g(n)$ and the vocal tract system, respectively.

This method is based on the minimization in the frequency domain, while the conventional methods are based on the minimization in the time domain.

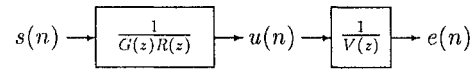


Fig.2 Analysis model for the proposed method.

3.2 Vocal Tract Analysis

After the speech signal is inverse-filtered through the glottal source inverse filter, the output signal($u(n)$) from the filter is analyzed using a conventional analysis method. Any speech analysis method can be used in this process. Therefore, the vocal tract filter stability is assured using the analysis method which assures the vocal tract filter stability, such as auto-correlation method and MEM method. Further, it is not necessary to determine the glottal wave position, since the glottal source inverse filter removes glottal source characteristics in the frequency domain. Therefore, the vocal tract parameters are calculated by solving only one equation corresponding to one assumed glottal wave and the computation amount can be significantly reduced. Thus, the method can solve two problems in the conventional method. In addition, the method does not require pre-emphasis.

3.3 Glottal Source Analysis

Simultaneous estimation of glottal and vocal tract parameters is a non-linear problem, since glottal parameters are non-linear parameters. A large amount of computation is required to reach the convergence of the solution in the iterative method. For that reason, in the proposed method, the glottal parameters are estimated by selecting an optimal glottal wave among a finite set of glottal waves, called the glottal codebook, instead of solving the non-linear equation. A set of glottal parameters is obtained by minimizing Eq.(2). The method leads to the linear problem. Further, the method has a capability to apply the further complex model, such as simultaneous use of different glottal wave models.

3.4 An Analysis Procedure

The analysis is carried out pitch synchronously. An analysis procedure of the proposed method is as follows:

1. A glottal code is selected from the glottal codebook in each pitch period.
2. The speech signal is filtered through the glottal source inverse filter, whose characteristics is determined by the selected glottal code.
3. The inverse filtered signal is analyzed using the conventional speech analysis method such as LPC analysis, which assures the vocal filter stability.

4. Procedures 2 and 3 are iterated for each of the selected glottal code.

5. Finally, the optimal glottal code is decided, which minimize Eq.(2).

4 EXPERIMENTS

4.1 Glottal Source Model

Several parametric models have already been proposed[3-6,8-10]. In this experiment, a time-domain six-parameter glottal source model is adopted[10]. This model consists of four phases. The differential waveform $\dot{g}(n)$ is expressed as follows.

[Phase1 : opening phase : $0 \leq n < N_1$]
[Phase2 : closing phase : $N_1 \leq n < P_3$]

$$\dot{g}(n) = \frac{G_1}{\sin\frac{\pi}{N_1}} r_1^n \sin\frac{(n+1)\pi}{N_1} \tag{3}$$

[Phase3 : close phase : $P_3 \leq n < P_4$]

$$\dot{g}(n) = G_3 \cdot r_3^n \cos\left(\frac{n\cdot\pi}{N_3} + \gamma + \delta\right) \tag{4}$$

$$G_3 = \left[\left(\frac{A}{\tan\frac{\pi}{N_3}} + \frac{B}{r_3 \sin\frac{\pi}{N_3}}\right)^2 + \left(\frac{A}{\sin\frac{\pi}{N_3}}\right)^2\right]^{\frac{1}{2}} \cdot \left[\left(\frac{\pi}{N_3}\right)^2 + (\log r_3)^2\right]^{\frac{1}{2}} \tag{5}$$

$$A = 2\cdot r_3 \cos\frac{\pi}{N_3} \cdot g(P_3-1) - r_3^2 \cdot g(P_3-2) \tag{6}$$

$$B = -r_3^2 \cdot g(P_3-1) \tag{7}$$

$$\tan\delta = \frac{\pi}{N_3 \log r_3} \tag{8}$$

$$\tan\gamma = \frac{A}{A\cos\frac{\pi}{N_3} + \frac{B}{r_3}} \tag{9}$$

[Phase4 : perfect close phase : $P_4 \leq n <$ pitch period]

$$\dot{g}(n) = 0 \tag{10}$$

This model has six parameters : three timing parameters (N_1, N_3, P_3) , two slope parameters(r_1, r_3) and an amplitude parameter(G_1). The shape of the glottal waveform and glottal source spectrum for this model can be easily controlled. Phases 3 and 4 correspond to the glottal closure. Phase 4 corresponds to a perfect glottal closure, where the glottal wave amplitude is zero. The length of Phase 1 and 2, where is glottal open period, is expressed by parameter P_3, the boundary point between Phase 2 and Phase 3. In Phase 3, the glottal waveform has negative amplitude, which corresponds to lowering the vocal cords followed by glottal closure, such as in the Fujisaki-Ljungqvist model[5]. The start point P_4 for Phase 4 is given by the first sample, where the glottal waveform amplitude is to be positive. The lip radiation is modeled by a differential filter as follows:

$$R(z) = 1 - z^{-1} \tag{11}$$

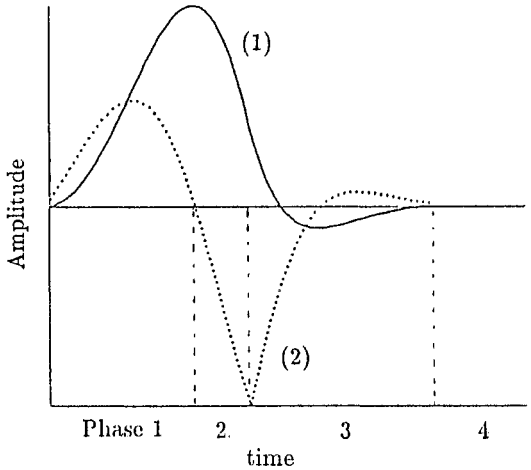
4.2 Analysis Experiment

Speech analysis-synthesis experiment was carried using the proposed analysis method. A speech sample tered by an adult male was used. Auto-correlation met was used for the vocal tract filter analysis. A 256 tap F filter was adopted to realize the glottal source inverse filt Table 1 summarizes analysis conditions.

Figure 4 shows estimated vocal tract and glot source spectrum characteristics during one pitch period a Japanese vowel /i/. Figure 5 shows the estimated p trajectories, which were obtained by solving the equat for the denominator polynomial of the estimated all-p system $(A(z))$, using the proposed method and the conv tional glottal LPC method. The glottal LPC method v carried out using the same analysis conditions in table 1 a the same glottal codebook used in the proposed method, a it determined the position of the glottal wave for the 40 sa ple period, whose center point was LPC residual peak sa ple. In the glottal LPC method, both the speech signal a the glottal wave was pre-emphasized by using a differen filter$(1-z^{-1})$. Figure 4 shows that the proposed method represent the speech spectrum well, since the estimated sp trum envelope conforms well to the DFT speech spectrum. can be seen from Figure 5 that the proposed method can timate more smooth pole trajectories than the conventio method. The conventional glottal LPC method fails to ac rately estimate the low frequency poles and the bandwi of the low frequency pole is very narrow due to pitch-rela bias. By using the proposed analysis method, since smo pole trajectories are obtained, poles between the adjac speech units can be easily interpolated. Further, synth speech was natural and intelligible. Therefore, it is expec that the proposed analysis is suitable for applying to spe synthesis by rule.

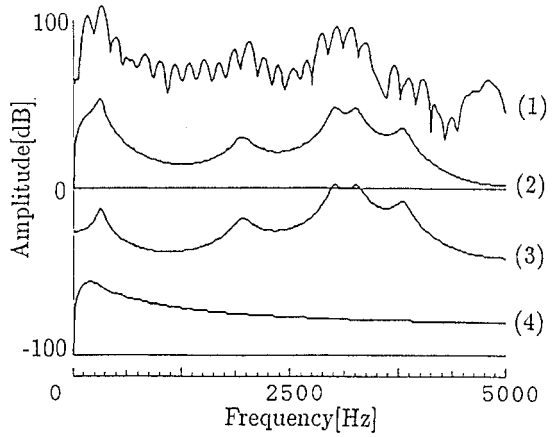
Table 1 Analysis condition

Analysis order	10
Analysis length	20[msec]
Analysis window	Hanning
Sampling rate	10[kHz]



(1) The glottal waveform $g(n)$
(2) The differential glottal waveform $\dot{g}(n)$

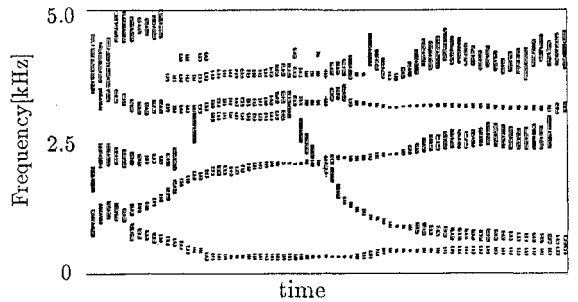
Fig.3 Six parameter glottal source model.



1)DFT spectrum
2)Estimated spectrum envelope
3)Estimated vocal tract characteristics
4)Differential glottal wave spectrum

Fig.4 Estimated vocal tract and glottal source spectra by the proposed method for Japanese vowel /i/.

(1)Glottal LPC method



(2)The proposed method

Fig.5 Estimated pole trajectories for Japanese vowel /aio/.

5 CONCLUSION

A new speech analysis method based on the glottal source model was proposed. The proposed method requires less amount of computation than the conventional method and assures vocal tract filter stability, since the method uses glottal source and vocal tract inverse filterings in the frequency domain and model parameters are calculated so as to minimize the inverse filtered error. The method can separate the voice source from the vocal tract characteristics, and can represent the speech spectrum more accurately than the conventional LPC method. The method can estimate more smooth pole trajectories than the conventional glottal LPC method. Accordingly, the poles between adjacent speech units can be easily interpolated. Further, synthetic speech is natural and intelligible. Thus, the proposed method is suitable for the rule-based speech synthesis system.

Acknowledgment

The authors wish to thank T.Watanabe and K.Ozawa for valuable comments for carrying out this study.

References

[1] T.Hirokawa, K.Hakoda, *"A Pitch Control Method for Waveform Concatenating Synthesis,"* Spring meeting of Acoustical Society of Japan,pp.191-192,March,1990(in Japanese).

[2] K.Iwata et.al., *"A Speech Synthesis System Using Pitch Controlled Residual Wave Excitation,"*Fall meeting of Acoustical Society of Japan,pp.183-184,October,1988(in Japanese).

[3] D.H.Klatt, *"View of text-to-speech conversion for English,"* JASA,82(3),pp.737-793,September,1987.

[4] G.Fant et.al., *"A Four-Parameter Model of Glottal Flow,"* KTH,STL-QPSR 4/85,pp.1-13,January,1986.

[5] H.Fujisaki et al., *"Comparative Evaluation of an ARMA Analysis Method Based on Modeling of the Glottal Source,"* Spring meeting of Acoustical Society of Japan,pp.137-138,March,1987.

[6] P.Hedelin, *"High Quality Glottal LPC-Vocoding,"* Proc.ICASSP-86,9.9.1.pp.465-468,1986.

[7] D.Y.Wong et.al., *"Least Squares Glottal Inverse Filtering from the Acoustic Speech Waveform,"* IEEE,Trans on ASSP-27,No.4,pp.350-355,1979.

[8] A.E.Rosenberg, *"Effect of Glottal Pulse Shape on the Quality of Natural Vowels,"* JASA,49,No.2,1971.

[9] T.V.Ananthapadmanabha, *"Acoustic Analysis of Voice Source Dynamics,"* KTH,STL-QPSR,No.2-3,pp.1-24,1984.

[10] Y.Mitome, *"A Speech Synthesis Model for Rule-based Synthesis,"* Spring meeting of Acoustical society of Japan,pp.189-190,March,1989(in Japanese).

[11] K.Funaki,Y.Mitome, *"Speech Analysis Synthesis Using Glottal Wave Model,"* Fall meeting of Acoustical society of Japan,pp.213-214,October,1989(in Japanese).