1999년도 한국음향학회 학술발표대회 논문집 제18권 제1(s)호

Split Model Speech Analysis Techniques for Wideband Speech Signal

*YoungHo Park, ***MyungKyu Ham, **KwangBock You, ***MyungJin Bae

*Dept. of Electronics Eng., Soongsil Univ. Seoul 156-743, Korea
**Dept. of Electrical and Computer Eng., Stevens Institute of Technology NJ 07030, USA
***Dept. of Telecomm Eng., Soongsil Univ. Seoul 156-743, Korea
e-mail : yhpark@assp.soongsil.ac.kr

ABSTRACT

In this paper, The Split Model Analysis Algorithm, which can generate the wideband speech signal from the spectral information of narrowband signal, is developed. The Split Model Analysis Algorithm deals with the separation of the 10th order LPC model into five cascade-connected 2nd order model. The use of the less complex 2nd order models allows for the exclusion of the complicated nonlinear relationships between model parameters and all the poles of the LPC model. The relationships between the model parameters and its corresponding analog poles is proved and applied to each 2nd order model. The wideband speech signal is obtained by changing only the sampling rate.

1. Introduction

In the LPC model, a speech Formant is characterized by a pole frequency and its bandwidth of the transfer function of the synthesis filter. However, it is difficult to synthesize a speech Formant structure by just pole modification. The major problem is pole interaction[1]. The objective of our enhancement algorithm is to reconstruct the wideband speech signal from the information contained in the narrowband one. From the narrowband speech signal with f_n sampling rate, it is to reconstruct the wideband speech signal with sampling rate of $f_w = 2f_n$. If a pole of the LPC model in normalized form can be expressed as $Me^{-j\omega}$ when the sampling rate is f_n and there is no aliasing, then the pole can be expressed as $Me^{-j\omega/2}$ when the sampling rate is f_w. Model parameters such as LPC coefficients, however, are products of highly nonlinear interactions of all the poles of the model. Hence the change of sampling rate results in a set of complicated relationships among model parameters unless the model is of second order.

Therefore, we split the 10th order LPC model, which is often used in speech analysis, into five 2nd order models in cascade form. Since each 2nd order model has only a pair of complex conjugate poles, the pole interaction does not exist in the model parameters. Two peak values, one of which must be located in the low-band (0-4kHz) and the other peak in the high-band (4-8kHz), exist in the spectrum of 2nd order model.

The purpose of speech enhancement algorithm is to reduce the background noise, improve speech quality, or suppress channel or speaker interference. Our new enhancement algorithm addresses the speech reproduction issue. In the following section of this Chapter, our states the important ideas and contribution of this study. Chapter 3 presents the Split Model Analysis with simulation results. Final discussion and possible future works are described in chapter 4.

2. Split Model Analysis

2.1 The Relationships between AR Parameters α_k and The Continuous Poles s_k

Consider a real wide-sense stationary(WSS) process $X_a(t)$ of finite order as the response of a minimum-phase system $H_a(s)$ with input a white-noise process v(t)[2]. The power spectrum of this system can be expressed as

$$S_a(s) = \sigma_v^2 H_a(s) H_a(-s) \tag{1}$$
$$\text{where} \quad \sigma_v^2 = E\{v(t)v(t)\}$$

Assuming that the poles s_k of $H_a(s)$ are simple, we expand $S_a(s)$ into partial fractions:

$$S_a(s) = \sigma_v^2 \left\{ \sum_{k=1}^{N} \frac{C_k}{s - s_k} + \sum_{k=1}^{N} \frac{C_k}{-s - s_k} \right\} \tag{2}$$

Where N represents the order of $S_a(s)$ and the coefficients C_k can be determined from the initial values theorem. On the other hand, the transfer function of the AR process of N^{th} order can be expressed as

$$H(z) = \frac{G}{1 + \sum_{k=1}^{N} \alpha_k Z^{-k}} \tag{3}$$

Where the coefficients $\{\alpha_k\}$ are real and G is a constant. The power spectrum for this system can be expressed as

$$S(z) = \sigma_v^2 H(z) H(z^{-1}) \tag{4}$$

and

$$\begin{aligned} S(z) &= \sigma_v^2 \sum_{k=1}^{N} C_k \frac{1 - e^{2 s_k T}}{(1 - e^{s_k T} z)(1 - e^{s_k T} z^{-1})} \\ &= \sigma_v^2 \frac{N(z)}{\prod_{k=1}^{N} (z - e^{s_k T}) \prod_{k=1}^{N} (z^{-1} - e^{s_k T})} \\ &= \sigma_v^2 \frac{G}{\sum_{k=0}^{N} \alpha_k z^{N-k}} \cdot \frac{G}{\sum_{k=0}^{N} \alpha_k z^{k-N}} \quad (\alpha_o = 1) \end{aligned} \tag{5}$$

Since denominators of Equation (5) can be regarded as the characteristic polynomials for this system, we conclude that

$$\sum_{k=0}^{N} \alpha_k z^{N-k} = \prod_{k=1}^{N} (z - e^{s_k T}) \tag{6}$$

This equation shows the relationships between the coefficients of AR process and its corresponding analog poles. From Equations (5) we know that the numerator N(z) of Eq.(5) is a constant, not a function of z. Since the Split Model Analysis is focused on the 2^{nd} order model, we can show that the numerator N(z) is a constant when the order is 2^{nd}. When N=2, the numerator N(z) in Equation (5) is expressed as

$$\begin{aligned} N(z) &= C_1(1 - e^{2 s_1 T})(1 + e^{2 s_2 T}) + C_2(1 - e^{2 s_2 T})(1 + e^{2 s_1 T}) \\ &\quad - \left\{ C_1(1 - e^{2 s_1 T}) e^{s_2 T} + C_2(1 - e^{2 s_2 T}) e^{s_1 T} \right\}(z + z^{-1}) \end{aligned} \tag{7}$$

First, examine the coefficients of $(z+z^{-1})$ term

$$\begin{aligned} & -\left\{ C_1(1 - e^{2 s_1 T}) e^{s_2 T} + C_2(1 - e^{2 s_2 T}) e^{s_1 T} \right\} \\ &= -e^{(s_1 + s_2)T} \left\{ C_1 e^{-s_1 T} + C_2 e^{-s_2 T} - (C_1 e^{s_1 T} + C_2 e^{s_2 T}) \right\} \\ &= -e^{(s_1 + s_2)T} \{R(-1) - R(1)\} \\ &= 0 \end{aligned} \tag{8}$$

$$\text{where} \quad R(1) = R(-1) = C_1 e^{s_1 T} + C_2 e^{s_2 T}.$$

We can calculate the constant term in Equation (7) using the same method:

$$\begin{aligned} & C_1(1 - e^{2 s_1 T})(1 + e^{2 s_2 T}) + C_2(1 - e^{2 s_2 T})(1 + e^{2 s_1 T}) \\ &= \{R(0) - R(2)\}(1 - e^{2(s_1 + s_2)T}) \end{aligned} \tag{9}$$
$$\text{where} \quad R(0) = C_1 + C_2, R(2) = C_1 e^{2 s_1 T} + C_2 e^{2 s_2 T}$$

Therefore, the power spectrum for 2^{nd} order model becomes:

$$S(z) = \sigma_v^2 \frac{\{R(0) - R(2)\}\left\{1 - e^{2(s_1 + s_2)T}\right\}}{\prod_{k=1}^{2} (z - e^{s_k T}) \prod_{k=1}^{2} (z^{-1} - e^{s_k T})} \tag{10}$$

2.2 Splitting the LPC Synthesis Filter

The wideband speech signal would not be reconstructed by the pole modification with changing the sampling rate due to the pole interaction of the transfer function of the LPC synthesis filter. Thus, the 10^{th} order LPC synthesis filter is split into five filters of 2^{nd} order to avoid the pole interpolation. The transfer function of the 10^{th} order LPC synthesis filter can be expressed as

$$H(z)=\frac{1}{A(z)}=\frac{1}{1+\sum_{k=1}^{10}\alpha_k z^{-k}} \tag{11}$$

$$=\frac{1}{\prod_{k=1}^{5}(1-p_k z^{-1})(1-p_k^* z^{-1})}$$

Where α_i's are prediction coefficients and p_i's and p_i^*'s are roots and complex conjugate roots of A(z), respectively. In section(2.1), we showed the relationships between the parameters of AR model and its corresponding analog poles. Therefore, we can describe the relationships between model parameters of the 2nd order filter and corresponding analog poles, i.e.,

$$\alpha_{k,1}=-(e^{s_k T}+e^{s_k^* T}) \tag{12}$$

$$\alpha_{k,2}=e^{(s_k+s_k^*)T}$$

Where the complex poles $s_k=\sigma_k+jw_k$. From Equation(12), then, we can calculate σ_k and jw_k.

$$\sigma_k=\frac{1}{2T}\ln(\alpha_{k,2}) \tag{13}$$

$$\omega_k=\frac{1}{T}\arccos\left(\frac{-\alpha_{k,1}}{2\sqrt{\alpha_{k,2}}}\right)$$

Now, applying new sampling interval T_0 to Equation (13), new model parameters with respect to T_0 are obtained

$$\beta_{k,1}=-(e^{s_k T_0}+e^{s_k^* T_0}) \tag{14}$$

$$\beta_{k,2}=e^{(s_k+s_k^*)T_0}$$

Therefore, we are able to obtain new transfer function of LPC synthesis filter for the wideband signal:

$$G_k(z)=\prod_{k=1}^{5}\frac{1}{1+\beta_{k,1}z^{-1}+\beta_{k,2}z^{-2}} \tag{15}$$

$$=\prod_{k=1}^{5}\frac{1}{(1-q_k z^{-1})(1-q_k^* z^{-1})}$$

Where q_k's and q^*_k's are new poles and complex conjugate poles for new sampling interval, respectively. It is important to notice that Equation(15) for the wideband speech signal model is obtained by changing only the sampling interval from T to T_0

3.Experimental and Result

Now, the Split Model Analysis of LPC synthesis filter which is described by Equation (11) through (15) is examined by the computer simulation. Our data which is sampled by 16kHz from the TIMIT speech data is /a/ sound of the world "hard" spoken by a female. We take 480 samples which is the length of 30ms as a processing frame. Then, we normalize the original 16kHz sampled speech data[3].

The Normalized speech signal is applied to the Split Model Analysis for generating the spectrum of 8kHz sampled data which is referred as the input data. The Split Model Analysis generates five 2nd order filter which are cascade connected.

Figure (1) shows the system block diagram for this simulation. For comparing convenience, the 16kHz sampled TIMIT speech data is referred as the original and the input data s(n) in Fig(1) is 8kHz sampled data which were down-sampled from the original data. In Fig(2), we show the spectral envelope for each signal. The spectral envelope of the original signal illustrates that it has only very small amount of important information between 4kHz and 8kHz. The spectral information of our reconstructed signal, described as dashed line, is compared and the spectrum of input signal is shown by the dotted line. The spectra of five 2nd order filters which are generated by the Split Model Analysis are shown in Fig(3) through(5).

4.Conclusion and Future Work

Our objective with this paper is the clear and through development of Split Model Analysis, which is based on the model-splitting concept. The Split Model Analysis deals with the separation of the 10th order LPC model into five cascade-connected 2nd order model. The use of the less complex 2nd order models allows for the exclusion of the complicated nonlinear relationships between model parameters and all the poles of the LPC model. As a result, we maintain simple relationships, even though the sampling rate is changed. By utilizing the spectrum features of a 2[nd] order model, we can calculate the spectrum of the wideband speech signal with simple computation. We can then reconstruct the wideband speech signal from the narrowband information. The increased simplicity of the Split Model Analysis, in our perspective, is intended to be a suggestion towards a new direction in the speech signal processing. In application field, The Split Model Analysis can be applied to video telephony and video conferencing system. Also, our studies pertaining to the Split Model Analysis, allows for the reduction of the number of bits to allocation to the input speech signal and furthermore lessens the load of computation these systems face.

Reference

[1] H.Mizuno, M.Abe and T.Hirokawa,"Waveform-based Speech Synthesis Approach with a Formant Frequency Modification", Proc. ICASSP'93, pp. II.195-198,1993

[2] S.Haykin, Adaptive Filter Theory, 2[nd] Edition, Prentice-Hall Inc., 1991

[3] A. Leon-Garcia, Probability and Random Process for Electrical Engineering, Addison-Wesley Publishing Company,1989

[4] TIA/EIA/IS-641, " TDMA Cellular/PCS- Radio Interface- Enhanced Full-Rate Speech Codec", Telecommunication Industry Association, Standards and Technology Dept., May,1996.

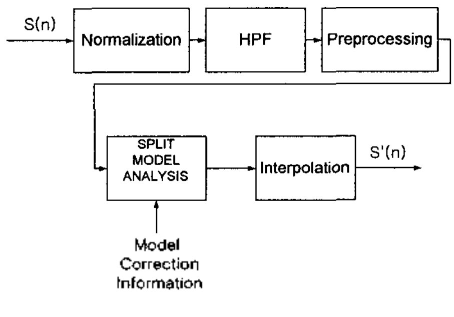

Fig 1. The Block diagram of the system

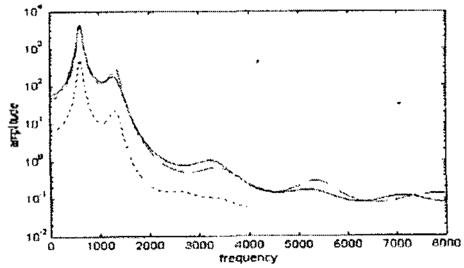

Fig 2. Spectral envelopes: '-' for 16kHz sampled original data, '— —'for 16kHz sampled reconstructed data and '- -'for 8kHz sampled input data.

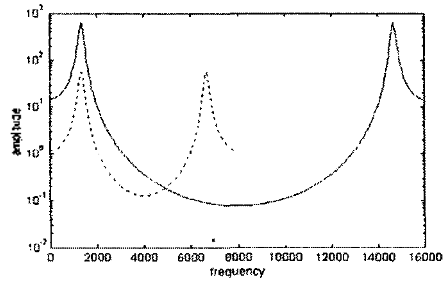

Fig 3. Spectral envelopes for the third 2[nd] order filter.

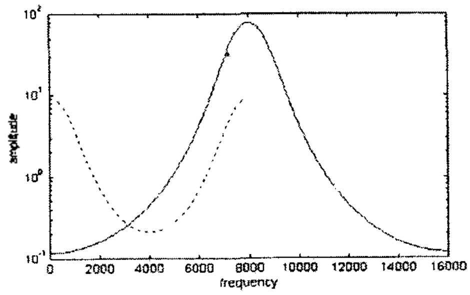

Fig 4. Spectral envelopes for the fourth 2[nd] order filter.

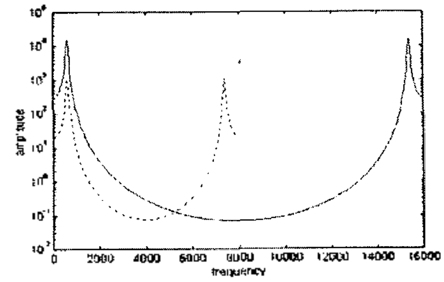

Fig 5. Spectral envelopes for the fifth 2[nd] order filter.