Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 63–64 (2014) 70–83

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Extraction of formant bandwidths using properties of group delay functions

Anand Joseph Xavier Medabalimi [a,*], Guruprasad Seshadri [b], Yegnanarayana Bayya [a]

[a] *International Institute of Information Technology, Hyderabad 500032, India*
[b] *Innovation Labs., TATA Consultancy Services, Bangalore 560066, India*

Received 27 September 2013; received in revised form 6 March 2014; accepted 21 April 2014
Available online 9 May 2014

Abstract

Formant frequencies represent resonances of vocal tract system during the production of speech signals. Bandwidths associated with the formant frequencies are important parameters in analysis and synthesis of speech signals. In this paper, a method is proposed to extract the bandwidths associated with formant frequencies, by analysing short segments (2–3 ms) of speech signal. The method is based on two important properties of group delay function (GDF): (a) The GDF exhibits prominent peaks at resonant frequencies and (b) the influence of one resonant frequency on other resonances is negligible in GDF. The accuracy of the method is demonstrated for synthetic signals generated using all-pole filters. The method is evaluated by extracting bandwidths of synthetic signals in closed phase and open phase regions within a pitch period. The accuracy of the proposed method is also compared with that of two other methods, one based on linear prediction analysis of speech signals, and another based on filterbank arrays for obtaining amplitude envelopes and instantaneous frequency signals. Results indicate that the method based on the properties of GDF is suitable for accurate extraction of formant bandwidths, even from short segments of speech signal within a pitch period.

Keywords: Formant frequency; Bandwidth; Group delay function; Short segments; Closed phase; Open phase

1. Introduction

During the production of speech, the nature of speech sounds depends on the time-varying characteristics of the source of excitation and those of the vocal tract system. Formants are resonances of the vocal tract system, and they represent important sound-specific and speaker-specific information (Fant, 1960). Bandwidths associated with the formant frequencies are useful parameters in the analysis and synthesis of speech signals. Accurate extraction of formant bandwidths from speech signals is a difficult task, since the formant frequencies and their bandwidths vary across pitch periods. Formant frequencies and their bandwidths vary even within a pitch period, from closed phase of glottis to the open phase. This is due to decoupling of trachea and vocal tract during the closed phase of glottis, and coupling of source-tract during the open phase of glottis. These issues necessitate analysis of short segments of speech (typically less than a pitch period) for extraction of formant bandwidths.

Some methods proposed in the literature for the extraction of formant bandwidths use model-based approaches for representation of speech signal. An approach based on AM–FM modeling of speech signal was proposed by Cohen et al. (1992), and an expression for formant bandwidth was obtained in terms of the parameters of the model. Potamianos and Maragos (1995) employed a

* Corresponding author.
E-mail addresses: anandjm@research.iiit.ac.in (A.J.X. Medabalimi), guruprasad.seshadri@gmail.com (G. Seshadri), yegna@iiit.ac.in (Y. Bayya).

http://dx.doi.org/10.1016/j.specom.2014.04.006
0167-6393/

bank of Gabor bandpass filters to decompose the speech signal, and the signal in each band was demodulated to obtain amplitude envelope and instantaneous frequency signals. Bandwidth estimates were obtained from the instantaneous frequency signals. An exponentially weighted autoregressive (EWAR) spectral model was proposed to extract the bandwidths of formant frequencies (Zheng and Hasegawa-Johnson, 2003). A method called clustered line-spectrum modeling was proposed to decompose the speech signal into three dominant resonant oscillations with nearly exponentially decaying envelopes (Yasojima et al., 2006). The bandwidths were estimated from the decaying constants of the resonant frequencies. In the above cases, the accuracy of extraction of bandwidths depends on the fitness of the models, and on the accuracy of estimation of model parameters from the speech signal. These methods typically use more than one pitch cycle of speech signal, a duration over which the bandwidths of formants tend to vary. Hence, there is need for methods to extract bandwidths from short segments (2–3 ms in duration) of voiced speech signals.

Linear prediction (LP) analysis is commonly used for extracting formant frequencies (Makhoul, 1975). The estimation of autoregressive parameters in LP analysis is based on an error minimization criterion. Reddy and Swamy (1984) proposed a method to extract bandwidths of formants, by observing the phase slope of the z-transform around the poles obtained using LP analysis of speech signal. However, the accuracy of formant frequencies and bandwidths depends on the choice of order of LP analysis. Also, the error minimization criterion in LP analysis focuses on matching spectral peaks, and bandwidth is only an additional outcome of the process.

In this paper, we propose a method for extraction of formant bandwidths by exploiting the properties of phase of Fourier transform. The method assumes that the speech signal in voiced regions can be modeled as the output of an all-pole filter. The key idea is based on the evaluation of group delay function at the resonant frequencies. In Section 2, some important properties of phase response of all-pole systems are revisited. These properties are exploited in Section 3, which describes the analytical basis for the proposed method. This section also examines the effectiveness of bandwidth extraction for all-pole systems. The method is evaluated for the case of synthetic speech signals in Section 4. Accuracy of the method is compared with two other methods of bandwidth extraction, and results are discussed. Conclusions are given in Section 5.

2. Properties of phase response of all-pole systems

In this section, we summarize the observations reported by Yegnanarayana (1978) on the significance of processing phase response and its derivative for extraction of formant frequencies from speech signals. These observations provide a background for the method proposed in Section 3.1, for extraction of formant bandwidths from discrete-time speech signals.

Let us consider a cascade of M resonators. The frequency response of the i^{th} resonator is given by

$$H_i(\omega) = \frac{1}{(j\omega - (\alpha_i + j\beta_i))(j\omega - (\alpha_i - j\beta_i))}, \tag{1}$$

where $\alpha_i \pm j\beta_i$ is the complex pair of poles of the i^{th} resonator, ω is the analog angular frequency and $j = \sqrt{-1}$. The expression is simplified as

$$H_i(\omega) = \frac{1}{\alpha_i^2 + \beta_i^2 - \omega^2 - 2j\omega\alpha_i}. \tag{2}$$

The squared magnitude response of the i^{th} resonator is given by

$$|H_i(\omega)|^2 = \frac{1}{(\alpha_i^2 + \beta_i^2 - \omega^2)^2 + 4\omega^2\alpha_i^2}. \tag{3}$$

The squared magnitude response of the overall filter, i.e., the cascade of M resonators, is given by

$$|H(\omega)|^2 = \prod_{i=1}^{M} |H_i(\omega)|^2. \tag{4}$$

The phase response of the i^{th} resonator is given by

$$\Theta_i(\omega) = \tan^{-1}\left(\frac{2\omega\alpha_i}{\alpha_i^2 + \beta_i^2 - \omega^2}\right). \tag{5}$$

Group delay function of the i^{th} resonator, which is the negative derivative of the corresponding phase response, is given by Yegnanarayana (1978)

$$G_i(\omega) = -\frac{2\alpha_i(\alpha_i^2 + \beta_i^2 + \omega^2)}{(\alpha_i^2 + \beta_i^2 - \omega^2)^2 + 4\omega^2\alpha_i^2}. \tag{6}$$

The group delay function of the overall filter is given by

$$G(\omega) = \sum_{i=1}^{M} G_i(\omega). \tag{7}$$

It can be verified that the magnitude response $|H_i(\omega)|^2$ (Eq. (3)) has a peak at $\omega^2 = \beta_i^2 - \alpha_i^2$, and a half power bandwidth of α_i. For sharp resonant peaks in the magnitude response, $\beta_i^2 \gg \alpha_i^2$. We now note the following properties of $G_i(\omega)$ (Yegnanarayana, 1978).

(a) From Eq. (3) and (6)

$$G_i(\omega) = -2\alpha_i(\alpha_i^2 + \beta_i^2 + \omega^2)|H_i(\omega)|^2. \tag{8}$$

For $\beta_i^2 \gg \alpha_i^2$, the group delay function $G_i(\omega)$ can be approximated around the resonant frequency $\omega^2 = \beta_i^2 - \alpha_i^2$ as follows:

$$G_i(\omega) = K_i|H_i(\omega)|^2, \tag{9}$$

where K_i is a constant. It is to be noted that $G_i(\omega)$ too has a peak near $\omega^2 = \beta_i^2 - \alpha_i^2$.