Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 81 (2016) 42–53
www.elsevier.com/locate/specom

An analysis of the high resolution property of group delay function with applications to audio signal processing

Jilt Sebastian[a,*], Manoj Kumar P. A.[b], Hema A. Murthy[c]

[a] Computer Science and Engineering, Indian Institute of Technology Madras, India
[b] Electrical Engineering Department, University of Southern California, United States
[c] Computer Science and Engineering, Indian Institute of Technology Madras, India

Received 1 July 2015; received in revised form 14 November 2015; accepted 26 December 2015
Available online 12 April 2016

Abstract

This paper provides a new insight into the high resolution property of the negative derivative of the phase response of a system. Group delay functions have been proposed and applied successfully as an alternative to conventional magnitude spectrum based applications in speech and music processing. One of the reasons claimed for its superior performance is the high spectral resolution. Most of the existing work use empirical analysis to show this property. In this paper, we show mathematically that for a single resonator, the ratio of the value of the peak in the magnitude spectrum to the value at a frequency that is n dB below the peak, is always much lower than the ratio of that of the minimum phase group delay spectrum. The results are extended for multiple resonators using numerical analyses. The theoretical results are reinforced using three applications, namely, pitch estimation, formant estimation and onset detection. The average deviation from the location of the pitch value/formant value/musical onset is about 53% lower than that of similar techniques that use the magnitude spectrum of the signal.

Keywords: Group delay; Spectral resolution; Pitch estimation; Formant estimation; Onset detection.

1. Introduction

In the frequency representation of speech signals, information is encoded in both magnitude and phase. The importance of the phase spectrum in speech technology has been established only in the last few decades. Intelligibility tests using phase only reconstruction in Alsteris and Paliwal (2007) indicate that temporal aspects of speech are encoded better in the phase compared to that of the magnitude. Shi et al. (2006) showed under various signal-to-noise ratios that having random phases for each frequency significantly altered the recognition rate as compared to actual (and reconstructed) phase. In Hegde et al. (2004), Rajesh M Hegde (2005) and Bozkurt and Couvreur (2005), the phase spectrum has been processed (via the group delay spectrum) for applications in speaker and speech recognition.

The negative frequency derivative of phase, commonly known as group delay function, has been proposed as an alternative to magnitude spectrum for segmentation (hybrid segmentation) and feature extraction tasks (ASR, TTS) in speech and music processing (Shanmugam and Murthy (2014b), Rajan and Murthy (2013a), Lakshmi and Murthy (2008), Rasipuram et al. (2008)). In all applications, the high spectral resolution of group delay functions is argued to contribute directly or indirectly to the superior performance. In syllable segmentation, spectral resolution enables accurate peak/valley detection while in pitch and formant estimation, resonances and anti-resonances are better emphasized by spectral resolution.

The high resolution property, although stated and acknowledged in most of the earlier applications involving group delay and derived functions, has not been well explained or proven. In earlier efforts towards justification, the group

* Corresponding author. Tel.: +91 9962876798.
E-mail addresses: jiltsebastian@gmail.com, jilt18@gmail.com (J. Sebastian), manojpamk@gmail.com (M. Kumar P. A.), hema@cse.iitm.ac.in (H.A. Murthy).

http://dx.doi.org/10.1016/j.specom.2015.12.008
0167-6393/

delay function was studied in the vicinity of resonances. In (Yegnanarayana, 1979), where formant estimation was attempted from the linear prediction phase spectra, a cascade of resonators was considered. For a constrained location of the poles, it was shown that the squared magnitude behavior of the group delay function around the resonance leads to its high resolution property. The magnitude spectrum for a cascade of N poles ($\alpha_i \pm \beta_i$ where $1 \leq i \leq N$) is given by:

$$|H(\omega)|^2 = \prod_{i=1}^{N} \frac{1}{(\alpha_i^2 + \beta_i^2 - \omega^2)^2 + 4\omega^2\alpha_i^2)} \tag{1}$$

The corresponding group delay spectrum is:

$$\theta'(\omega) = \sum_{i=1}^{N} \frac{2\alpha_i(\alpha_i^2 + \beta_i^2 - \omega^2)}{(\alpha_i^2 + \beta_i^2 - \omega^2)^2 + 4\omega^2\alpha_i^2)} \tag{2}$$

For $\beta_i^2 \gg \alpha_i^2$, the group delay can be approximated as

$$\theta'(\omega) \simeq \sum_{i=1}^{N} \frac{K_i}{(\alpha_i^2 + \beta_i^2 - \omega^2)^2 + 4\omega^2\alpha_i^2)}$$

or

$$\theta'(\omega) \simeq \sum_{i=1}^{N} K_i |H(\omega)|^2 \tag{3}$$

where K_i is a constant. Hence, most of the energy in the group delay domain is argued to be concentrated around the resonator, thus enabling better formant estimation. Clearly, the proof places a significant constraint on the pole, that it must be close to the imaginary axis and have a small bandwidth. However, this assumption is not very realistic since the most important formants (e.g. for formant tracking) are often close to the real axis and with small imaginary component.

Later (Bellur and Murthy, 2013) considered a parallel connection of resonators while working on pitch histograms that resembled a non-constant Q factor for each of the peaks. By considering an example of two resonators in parallel, it was shown that the group delay response was approximated by the squared magnitude spectrum around the peaks. Both (Yegnanarayana, 1979) and (Bellur and Murthy, 2013) consider only the region around the peaks, and present comparativeness rather than analysis proof for peakedness.

In this paper, a proof is presented for the high resolution property. Section 2 considers the case of a single resonator system, and proves, without any constraints on the location of poles, that the group delay function always possesses a sharper peak in comparison to the magnitude spectrum. The strength of the group delay function at the n dB bandwidth of the magnitude spectrum is always shown to be lesser. Section 3 quantifies the resolution for multi-resonator systems using numerical computations and empirical measures such as kurtosis and spectral flatness. In Section 4, three applications are discussed that benefit from the high resolution property, and Section 5 draws comparisons between group delay (phase) based and magnitude based approaches through experimentation. Conclusions are presented in Section 6.

2. A single-resonator minimum-phase system - Theoretical approach

Consider a causal, discrete-time signal $x[n]$ with one pole whose location in the z-plane is given as $z_0 = re^{j\omega_0}$, or $z_0 = e^{-\sigma_0 + j\omega_0}$. σ_0 represents the bandwidth of the pole and ω_0 the angle with respect to the abscissa. The Z-transform of the above system is:

$$X(z) = \frac{1}{(z - z_0)(z - z_0^*)} \tag{4}$$

When evaluated at the unit circle:

$$X(\omega) = \frac{1}{(e^{j\omega} - e^{-\sigma_0 + j\omega_0})(e^{j\omega} - e^{-\sigma_0 - j\omega_0})} \tag{5}$$

The expression for the magnitude spectrum is given as:

$$|X(\omega)| = P \times Q \tag{6}$$

where

$$P = \frac{1}{\sqrt{1 + e^{-2\sigma_0} - 2e^{-\sigma_0}\cos(\omega - \omega_0)}} \tag{7}$$

$$Q = \frac{1}{\sqrt{1 + e^{-2\sigma_0} - 2e^{-\sigma_0}\cos(\omega + \omega_0)}} \tag{8}$$

Considering (7) alone, the maximum value, $\frac{1}{1-e^{-\sigma_0}}$, occurs at $\omega = \omega_0$. To compute the n dB bandwidth, we determine the angular frequency (ω_1) at which the magnitude spectrum falls to $\frac{1}{N}$ of its maximum value, i.e

$$\frac{1}{\sqrt{(1 + e^{-2\sigma_0} - 2e^{-\sigma_0}\cos(\omega_1 - \omega_0))}} = \frac{1}{N(1 - e^{-\sigma_0})} \tag{9}$$

Here, $N = 10^{\frac{n}{20}}$. Solving for ω_1,

$$\omega_1 = \omega_0 \pm \cos^{-1}\left(N^2 + \frac{1 - N^2}{2}(e^{\sigma_0} + e^{-\sigma_0})\right) \tag{10}$$

The n dB bandwidth is the interval with ω_0 at the center, and is given by

$$\omega_{ndB} = 2\cos^{-1}\left(N^2 + \frac{1 - N^2}{2}(e^{\sigma_0} + e^{-\sigma_0})\right) \tag{11}$$

We repeat this analysis for the group delay spectrum. The phase spectrum for the system defined by (5) is given by

$$\theta(\omega) = -\tan^{-1}\left(\frac{\sin(\omega) - e^{-\sigma_0}\sin(\omega_0)}{\cos(\omega) - e^{-\sigma_0}\cos(\omega_0)}\right) - \tan^{-1} \times \left(\frac{\sin(\omega) + e^{-\sigma_0}\sin(\omega_0)}{\cos(\omega) - e^{-\sigma_0}\cos(\omega_0)}\right) \tag{12}$$

The group delay is defined as the negative derivative of the phase spectrum and is given by

$$GD(\omega) = \frac{1 - e^{-\sigma_0}\cos(\omega - \omega_0)}{1 + e^{-2\sigma_0} - 2e^{-\sigma_0}\cos(\omega - \omega_0)} + \frac{1 - e^{-\sigma_0}\cos(\omega + \omega_0)}{1 + e^{-2\sigma_0} - 2e^{-\sigma_0}\cos(\omega + \omega_0)} \tag{13}$$

Differentiating the first term in (13) and equating to zero, we find that it displays the same abscissa and ordinate for