EUSIPCO 2013 1569745887

A NOVEL APPLICATION OF GROUP DELAY FUNCTION FOR IDENTIFYING TONIC IN CARNATIC MUSIC

Ashwin Bellur[1] *and Hema A Murthy*[2] *

[1] Department of Electrical Engineering,
[2] Department of Computer Science and Engineering,
Indian Institute of Technology, Madras, India - 600 036

ABSTRACT

In this work, we propose a novel application of the group delay function – to identify the tonic pitch in Carnatic music (one of the main sub genres of Indian classical music). Tonic pitch is the reference note chosen by a performer in Carnatic music. Automatic identification of the tonic pitch is essential in Carnatic music in order to perform data driven computational melodic analysis. Group delay functions, which have found numerous applications in speech processing are employed in this task to process pitch histograms to emphasize certain characteristics of Carnatic music that manifest in a pitch histogram, to aid in accurate tonic identification. In order to do so, the pitch histograms are first characterized as the squared magnitude response of a set of resonators in parallel that do not have constant Q. Some interesting properties of the group delay function of this magnitude response has then been illustrated and exploited to identify the tonic pitch. The proposed method is then tested by identifying tonic on a large varied database with an accuracy of 90.70%.

Index Terms— Group Delay, Parallel Resonators, Pitch, Histograms, Tonic

1. INTRODUCTION

The group delay function is defined as the negative derivative of the phase spectrum of a given signal. Using group delay functions, it has been shown that the phase spectrum (which generally appears to be noisy owing to its being wrapped) can be gainfully processed to extract useful information. Group delay functions [1] have been extensively studied in the context of speech processing, for both formant and pitch estimation [2].

Group delay based features have also found applications in speaker and speech recognition systems [3, 4, 5]. One of the flavors of group delay processing – the minimum phase group delay function derived from the magnitude spectrum of a speech signal has been used for formant extraction [2]. Similarities between magnitude spectrum and short term energy function has resulted in extensive work on segmentation of continuous speech into syllable-like units using minimum phase group delay functions [6]. Such processing methods have also found applications in synthesis too [7]. Almost all of these efforts, exploit the fact that the group delay response for a cascade of resonators, behaves like the squared magnitude response in the vicinity of the resonances, as detailed in [8]. The fact that the summative nature of the group delay spectrum offers better resolution and tracking of formants when compared to the magnitude spectrum have been fully explored in these works.

In this work, we propose another novel application of the group delay function. We first show that histograms constructed on the pitch extracted for an item in Carnatic music, owing to its unique features, can be characterized as the magnitude response of a system of resonators in parallel. We then illustrate some of the interesting properties of the minimum phase group delay spectrum vis a vis the magnitude spectrum for such a parallel setup. It is then shown that these properties of the group delay spectrum can be used to emphasize certain traits of the pitch histogram to identify tonic pitch.

To give perspective to the proposed approach, the task of tonic identification in Indian music is first discussed in section 2. Typical characteristics of pitch histograms are detailed in this section. Owing to these typical characteristics of histograms, in section 3, the pitch histogram is modeled as the squared magnitude response of a set of resonators in parallel. The advantages of processing such a model using group delay functions is illustrated using synthetic examples. In section 4, the procedure for processing actual pitch histograms using group delay function for tonic identification is discussed. In section 5, the performance of the proposed technique is evaluated. We conclude in section 6.

2. TONIC IDENTIFICATION IN CARNATIC MUSIC USING PITCH HISTOGRAMS

This work addresses the problem of tonic identification in Indian classical music. In Indian classical music, the tonic is the reference pitch with respect to which all the other notes

*This research was partly funded by the European Research Council under the European Unions Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

in a melody are defined. The tonic is decided by the lead performer, usually lead vocal, depending upon his/her vocal range. The lead performer generally uses a drone in the background, set to a particular frequency to establish the tonic. All other instruments that are part of the ensemble, also tune to the tonic or the reference pitch. This is in contrast to Western classical music where a movable tonic is used but a fixed frequency is employed as reference for tuning (usually *A4* at $440Hz$). There is no absolute fixed reference tuning frequency in Indian classical with tonic of lead performer serving as the reference.

In [9], an attempt was made to identify tonic by processing the drone using a multi pitch approach. Although the drone contains the tonic information, the drone is generally in the background and is seldom audible especially in yesteryear Carnatic Music performances. In this paper, the drone is not required to be present. The tonic is estimated by processing the melodic histograms for each item in Carnatic Music.

Given that Carnatic music is heterophonic in nature i.e all melodic sources render the same melody, in this work, only mono pitch is extracted. Yin [10] is used to extract pitch from a musical item. Figure 1 shows a typical histogram from $55Hz$ to $300Hz$ computed on the pitch extracted from a 3 minute excerpt of a Carnatic item, with a bin width of $1Hz$.

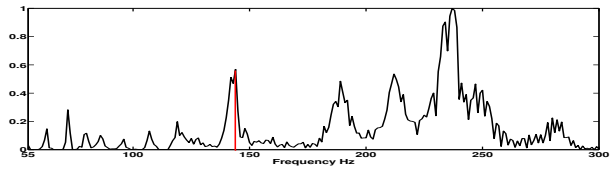

Fig. 1: Pitch histogram of a 3 minute Carnatic music item. The red stem is the tonic pitch value

To apply appropriate signal processing techniques, a large number of pitch histograms of Carnatic music were studied. The following observations were made:

- A peak is always seen at the tonic pitch value in the histogram, though not necessarily the most prominent. This is because, the drone (if present) is tuned to the tonic pitch, and percussion and accompanying instruments are also tuned to the tonic. Hence a peak is invariably observed at the tonic.

- The distribution of the pitch frequency appears to be continuous in nature, even though a melody is based on a scale of finite notes. Analyzing pitch histograms in Carnatic music, it was shown in [11] that tuning in Carnatic music tends towards the *just intonation* system, though with a continuous distribution in pitches. This apparent continuous distribution of pitches can be attributed to the inflected nature of notes [12], which is a fundamental characteristic of Carnatic music. There seldom are clear peaks representing all the notes of the melody at the expected theoretical positions and are sometimes difficult to resolve. This property of Carnatic music makes the peak picking process difficult.

- Another characteristic of Carnatic music is that notes that constitute a melody have dissimilar inflections, which manifest as peaks with varying bandwidth in the pitch histogram. It was observed that the tonic as well as the fifth (1.5 times tonic) are inflected less. The peaks corresponding to these notes have narrower bandwidths in the pitch histogram.

The above observations imply that in order to identify the tonic pitch value, it is essential to resolve and pick peaks from a pitch histogram that appears continuous. Given that the tonic will invariably have a prominent peak, further analysis of peaks and their bandwidths might suffice to identify tonic from pitch histograms. An effort in this direction was attempted in [13] using Gaussian Mixture Models to characterize histograms on a small dataset. The methods assumes peaks at expected theoretical frequencies of notes which is not necessarily the characteristic of Carnatic music [11]. In this work, we propose a knowledge based signal processing technique to process pitch histograms, in order to identify the tonic pitch.

3. PITCH HISTOGRAMS AS A SQUARED MAGNITUDE RESPONSE

In [6] it was shown that a positive function, short term energy in that case, can be treated as a square magnitude response, lending itself viable to group delay processing. In this work pitch histogram, a positive function, is processed using the group delay function. Similar to short term energy in [6], the pitch histogram can be assumed to be the squared magnitude $|H(e^{j\omega})|^2$ response of a system $H(e^{j\omega})$. At the same time, unlike the magnitude spectrum in speech or short term energy, where all formants/peaks are generally thought of as response constant Q filters from a cascade of resonators, some of the peaks in the pitch histogram can be characterized by both a large gain and a large bandwidth. This is primarily because the gain corresponds to the frequency of occurrence of a particular note in a given melody. To mimic the pitch histogram and to achieve a particular peak-gain at resonance, $H(e^{j\omega})$ can be thought of as the squared magnitude response of resonators connected in parallel with different gains, rather than a cascade of resonators:

$$H(z) = \sum_{k=1}^{M} \frac{\alpha_k}{(1 - d_k z^{-1})} \tag{1}$$

with $d_i = r_i e^{(j\omega_i)}$, r_i is the radius of the pole and ω_i is the angle of the pole in the Z plane.

To justify this model, we estimate the squared magnitude response of two resonators in parallel given by Equation 2.

$$H(z) = \frac{\alpha_1}{(1 - d_1 z^{-1})} + \frac{\alpha_2}{(1 - d_2 z^{-1})} \tag{2}$$

with poles at $d_1 = 0.9e^{(j\pi/4)}$, $d_2 = 0.7e^{(j\pi/3)}$ and different values of α_1 and α_2.

The angular frequency of the pole corresponds to the location of the peak, while the radius of the pole relates to the bandwidth of the pole. As the poles are close to each other, it can be seen in column 1 of Figure 2, that the two peaks are not resolved.

In this parallel setup, the width of the peak does not necessarily imply a decrease in gain. A gain term is included in Equation 2 to account for the height of the peak. A gain term can be accounted in the parallel representation of resonances. The Z-transform of the parallel connection as:

$$H(z) = (\alpha_1 + \alpha_2)\frac{1 - c_1 z^{-1}}{(1 - d_1 z^{-1})(1 - d_2 z^{-1})} \tag{3}$$

with $c_1 = \dfrac{\alpha_2 d_1 + \alpha_1 d_2}{\alpha_1 + \alpha_2}$.

The first two columns of Figure 2 show the squared magnitude spectrum and pole-zero plot respectively, for the system given by Equation 3. In this system α_1 is set to 1, while α_2 is varied. Since the model assumed is a parallel connection, zeroes are introduced as indicated in Equation 3. It can be seen that as α_2 increases, the zero moves towards the pole of the other resonator thus annihilating the pole of the second resonator.

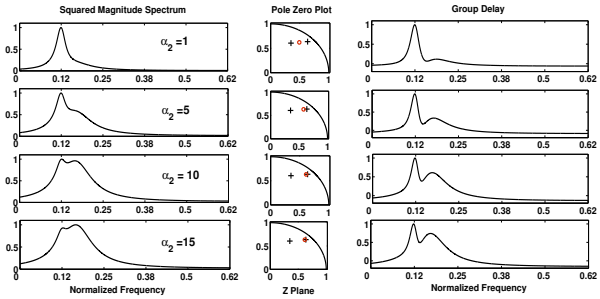

Fig. 2: Column 1 - Squared magnitude response, Column 2 - Pole zero plot, Column 3 - Group delay plot, for a two pole system with $\alpha_1 = 1$ and varying α_2

Pitch histograms exhibit similar behavior, i.e., peaks at various bin values can have varying heights and bandwidths. This primarily depends on the frequency of the note and the inflection that is associated with the note. Pitch histograms can therefore be characterized as a set of resonators in parallel with an appropriate choice of pole angles (ω_i), gains (α_i) and pole radii (r_i) respectively. Clearly, estimating position of peaks and their bandwidths even for the synthetic spectrum of Figure 2 is nontrivial. Given that pitch histograms in Carnatic music show similar traits, tonic identification from pitch histograms of Carnatic music is difficult. We conjecture that by processing such histograms using the group delay function will aid in tonic identification. Having shown that the histogram can be assumed to be the squared magnitude response of the system in Equation 3, we will now illustrate some of the properties of the group delay response of such a setup. These properties are later exploited to identify the tonic pitch.

The system in Equation 3 modeling a 2 peak histogram can be generalized as [14]

$$H(z) = G\frac{\prod_{k=1}^{M-1}(1 - c_k z^{-1})}{\prod_{k=1}^{M}(1 - d_k z^{-1})} \tag{4}$$

where M denotes the M peaks of a histogram and G is the gain. The squared magnitude spectrum of this system evaluated on the unit circle is given by:

$$|H(e^{j\omega})|^2 = (G)^2\frac{\prod_{k=1}^{M-1}(1 + |c_k|^2 - 2Re\{c_k e^{-j\omega}\})}{\prod_{k=1}^{M}(1 + |d_k|^2 - 2Re\{d_k e^{-j\omega}\})} \tag{5}$$

where Re implies the real component. For the system given in Equation 4, the group delay function is defined as the negative derivative of the continuous phase function of the system ($\arg[H(e^{j\omega})]$), i.e

$$\tau(\omega) = -\frac{d}{d\omega} arg[H(e^{(j\omega)})] \tag{6}$$

The group delay function of the system given in Equation 4 will take the form [14]

$$\begin{aligned}\tau(\omega) = &\sum_{k=1}^{M}\frac{|d_k|^2 - Re\{d_k e^{-j\omega}\}}{1 + |d_k|^2 - 2Re\{d_k e^{-j\omega}\}} \\ &- \sum_{k=1}^{M-1}\frac{|c_k|^2 - Re\{c_k e^{-j\omega}\}}{1 + |c_k|^2 - 2Re\{c_k e^{-j\omega}\}}\end{aligned} \tag{7}$$

For any pole with $d_i = r_i e^{j\omega_i}$, as $\omega \to \omega_i$, group delay function takes the form,

$$\frac{|d_i|^2 - Re\{d_i e^{-j\omega}\}}{1 + |d_i|^2 - 2Re\{d_i e^{-j\omega}\}} \approx \frac{K_i}{1 + |d_i|^2 - 2Re\{d_i e^{-j\omega}\}} \tag{8}$$

where K_i is some constant.

It can be seen from Equations 5, 7, and 8 that the group delay behaves like the squared magnitude response of the resonators at the resonance frequencies [8]. The additive property of the group delay is evident in the third column of Figure 2. While in the squared magnitude spectrum (column 1), the peaks (pole angles) are not resolvable for small values of α_2 and bandwidths (pole radius) are difficult to discern at high values of α_2, resolution of the peaks and bandwidth information are better preserved in the group delay extracted (Column 3 of Figure 2) .

To further illustrate the characteristics of the group delay function, the squared magnitude, pole-zero plot and group delay response for three resonators in parallel are shown in Figure 3. α_1 and α_3 are fixed at $\alpha_1 = \alpha_3 = 1$, with $d_1 = 0.9e^{(j\pi/4)}$, $d_2 = 0.7e^{(j\pi/3)}$ and $d_3 = 0.9e^{(j4\pi/3)}$. Square magnitude and group delay response is shown for various values of α_2. It is interesting to see in the group delay column, the peak due to the pole with a larger pole radius of 0.9 at $d3$ dominates even at large values of α_2. This ability of the group delay function to accentuate peaks with narrow bandwidths is indeed crucial for tonic identification.

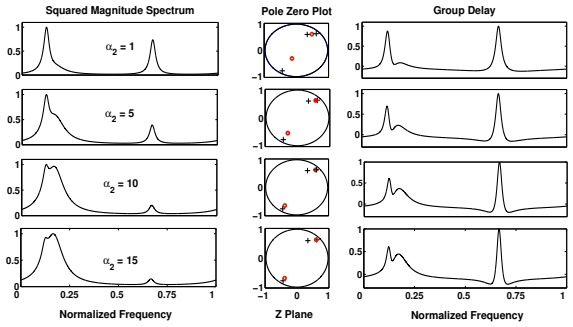

Fig. 3: Column 1 - Squared magnitude response, Column 2 - Pole zero plot, Column 3 - Group delay plot, for a three pole system with $\alpha_1 = \alpha_3 = 1$ and varying α_2

4. GROUP DELAY PROCESSING OF HISTOGRAMS

In the previous section, the advantages of group delay processing were illustrated through a synthetic example. In this section, extraction of group delay from a natural pitch histogram is described. The procedure is similar to the algorithm used for syllable segmentation in [6].

- In order to treat the pitch histogram as a squared magnitude spectrum, append to the pitch histogram, a symmetric histogram generated by lateral inversion about the Y-axis. If the original pitch histogram was made of N bins, it is now extended to 2N-1 bins, with its symmetric counterpart from bin N+1 to 2N-1. Let the symmetric histogram be $P[k]$.
- The causal portion of the inverse Fourier transform of a power function has minimum phase properties [15]. In order to extract the minimum phase signal, compute inverse discrete Fourier transform, $IDFT(P[k])$ and let the resultant sequence be $p[n]$.
- Extract the minimum phase signal by windowing the causal portion of $p[n]$ with a Hamming window of size N.
- If $c[n]$ is the windowed causal portion of $p[n]$, the minimum phase group delay $\tau[k]$ is estimated as given in Equation 9. Δ denotes first difference and arg denotes unwrapped phase.

$$\tau[k] = -\Delta arg[DFT(c[n])] \quad (9)$$

- $\tau[k]$ is henceforth referred to as the GD histogram.

Figure 4(b) shows the result of group delay processing the pitch histogram in Figure 4(a). Observe that the histogram is not only smoothened but also the peaks are sharper. Also here, peaks corresponding to that of the tonic and fifths (peaks with narrow bandwidth), are emphasized in the GD histogram. Methods to identify tonic are proposed in the following section utilizing these very features of group delay processing.

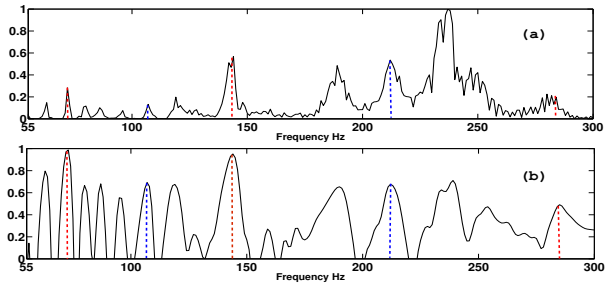

Fig. 4: Pitch and Group Delay Histograms (a) Pitch histogram of a 3 minute Carnatic music item. (b) The GD histogram. Red stems indicate the bin values of the tonic in the lower, middle and upper octaves. Blue stems indicate the position of the fifths in the middle and lower octaves

5. PERFORMANCE EVALUATION

In order to validate the proposed approach, tonic identification was performed on a large varied set of 344 Carnatic music excerpts. Excerpts are of 3 minute duration and pitch was extracted using Yin with a hop size of 0.01s. The pitch histogram was then computed with bin centers from $30Hz$ to $800Hz$, $1Hz$ apart. The ground truth was compiled by a professional musician. The tonic pitch values of the excerpts in the database were seen to range from $110Hz$ to $250Hz$. Two methods were attempted to identify the tonic using histograms and GD histograms. The tonic pitch as identified by these methods is deemed accurate if it is within 2 Hz range of the actual tonic pitch value.

1. Given that tonic pitch value is generally dominant in Carnatic music, the bin value of the tallest peak in the pitch histogram, between $110Hz$ to $250Hz$ is estimated as the tonic pitch value (Figure 4).
2. This method attempts to exploit the fact that the tonic pitch and the fifth in all of the octaves are less inflected compared to the other notes of the melody. The following procedure is used: peaks are first picked from the Histograms/GD histogram denoted by H. Now a vector G of the same dimension as H is constructed such that:
   For $i \in [30, 800]$, $G(i) = H(i)$ if bin i corresponds to a peak in the histogram H. $G(i) = 0$ for all remaining bins values. Each non zero value of G, corresponding to a peak in the histogram/GD histogram, is now treated as a candidate tonic.

   As mentioned in Section 2, in Carnatic music, for the actual tonic pitch value, say at bin f, one can expect peaks with narrow bandwidth at higher $(2f)$ and lower $(f/2)$ octave tonic pitch, and also at middle $(3f/2)$ and lower $(3f/4)$ octave fifths. This can be seen in Figure 4, where red stems indicate tonic and blue stems the fifths. To utilize this feature, for every candidate peak i, T is defined such that $T(i) = \sum_{k=-\delta}^{\delta} G(i/2 + k) +$

$G(3i/4+k)+G(i)+G(3i/2+k)+G(2i+k)$, with $\delta = 3$. For the correct candidate i.e $i = f$, $T(f)$ will be maximum for the GD histogram, given that each of the peaks that constitute T will be accentuated by the group delay function (Figure 4, b).

Table 1: Results comparing performance of Histograms and GD histograms on employing methods 1 and 2 to identify tonic pitch value

Method	Histogram	GD Histogram
Tallest Peak	72.67%	83.85%
Template Matching	84.01%	90.70%

It can be seen in Table 1 method 1, that the tonic pitch indeed features prominently in the music, being the tallest peak of the constructed histogram in around 73% of the cases. Due to further accentuation by the group delay function, the accuracy of method 1 improves to around 84% on using the GD histogram. In method 2, the ability of the group delay function to resolve peaks is first used to generate candidate tonic values that constitute G. The fact that the peaks corresponding to the fifth along with the tonic pitch is also emphasized is then utilized to achieve an accuracy of around 91% on using the GD histograms.

6. CONCLUSION

There have been many efforts to show the advantages of group delay processing, especially for extracting formants by assuming the system to be a cascade of resonators. In this work, a parallel setup of resonators has been used to characterize pitch histograms for the purpose of tonic identification in Carnatic music. By assuming pitch histograms to be squared magnitude response of resonators in parallel, characteristics of the group delay functions of such a setup, that can be favorably used to identify tonic has been highlighted. First, the ability of group delay function to resolve peaks and retain pole radii information for a set of parallel resonators has been illustrated. This property of group delay functions are shown to be crucial for emphasizing certain characteristic traits of Carnatic music that enable easy identification of tonic. The proposed approach was validated by testing on a large varied dataset.

7. REFERENCES

[1] B. Yegnanarayana, D. K. Saikia, and T. R. Krishan, "Significance of group delay functions in signal reconstruction from spectral magnitude or phase," *IEEE Trans. Acoustics Speech and Signal Processing*, vol. ASSP-32, no. 3, pp. 610–623, 1984.

[2] H. A. Murthy and B. Yegnanarayana, "Formant extraction from minimum phase group delay function," *Speech Communications*, vol. 10, pp. 209–221, August 1991.

[3] R. M. Hegde, H. A. Murthy, and V. R. R. Gadde, "Significance of the modified group delay features in speech recognition," *IEEE International Transactions on Audio,Speech and Language Processing*, vol. 15, pp. 190–202, January 2007.

[4] B. Bozkurt, L. Couvreur, and T. Dutoit, "Chirp group delay analysis of speech signals," *Speech Communication*, vol. 49, no. 3, pp. 159–176, March 2007.

[5] R. Padmanabhan and H. A. Murthy, "Dynamic selection of magnitude and phase based acoustic feature streams for speaker verification," in *Proceedings of European Conference on Signal Processing*, September 2009, pp. 1244–1248.

[6] T. Nagarajan, H. A. Murthy, and R. M. Hegde, "Segmentation of speech into syllable-like units," in *Proceedings of EUROSPEECH*, Geneva, Switzerland, September 2003, pp. 2893–2896.

[7] M. N. Rao, S. Thomas, T. Nagarajan, and H. A. Murthy, "Text-to-speech synthesis using syllable-like units," in *National Conference on Communication*, 2005, pp. 227–280.

[8] B. Yegnanarayana, "Formant extraction from linear prediction phase spectra," *Acoustical Society of America*, vol. 63, pp. 1638–1640, 1979.

[9] J. Salamon, S. Gulati, and X. Serra, "A multipitch approach to tonic tdentification in indian classical music," *In Proc. of ISMIR*, pp. 157–162, 2012.

[10] A. D. Cheveigne and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, pp. 111(4):1917–1930, 2002.

[11] J. Serra, G. K. Koduri, M. Miron, and X. Serra, "Tuning of sung indian classical music," *In Proc. of ISMIR*, pp. 157–162, 2011.

[12] A. Krishnaswamy, "Application of pitch tracking to south indian classical music." *In Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 557–560, 2003.

[13] H. G. Ranjani, S. Arthi, and T. V. Sreenivas, "Shadja, swara identification and raga verification in alapana using stochastic models," *In 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 29–32, 2011.

[14] A. V. Oppenheim and R. W. Schafer, *Discrete Time Signal Processing*. New Jersey: Prentice Hall, Inc, 1990.

[15] T. Nagarajan, V. K. Prasad, and H. A. Murthy, "Minimum phase signal derived from the root cepstrum," *IEE Electronics Letters*, vol. 39, pp. pp.941–942, June 2003.