ON ERRORS IN PITCH TRACKING OF MUSIC WITH TABLA ACCOMPANIMENT

Ashutosh Bapat[*], *Preeti Rao*[†], *Hari Sahasrabuddhe*[‡]

Indian Institute of Technology, Bombay, India
email: ashutosh@cse.iitb.ac.in

ABSTRACT

Pitch tracking is an important component of Automatic Music Transcription (AMT) systems and Query by Humming (QBH) based melody retrieval systems. While the detection of pitch (or fundamental frequency) is accomplished relatively easily for an individual voice, the presence of percussive accompaniment such as Tabla can greatly perturb the pitch tracker. With a view to build pitch detectors that are robust to percussive background, we present a study of the acoustic characteristics of common tabla strokes. An experimental investigation of the errors in autocorrelation function-based pitch detection is presented for a singing voice accompanied by tabla.

1. INTRODUCTION

Automatic Music Transcription (AMT) system is a system which converts a given audio file into the full musical score. A melody retrieval system [1] based on acoustic querying would allow a user to hum or sing a short fragment of a song into a microphone and then search and retrieve the best matched song from the database. Melody is defined by the time variation of the pitch (i.e. fundamental frequency [2]). Therefore both of these systems require pitch detector as their core block. Pitch detection is the process of estimating the pitch values at discrete intervals of time. The output of a pitch detector is typically a pitch contour i.e. plot of pitch versus time.

Over last two decades many Pitch Detection Algorithms (PDAs) have been developed [2]. In this paper we present experiments with very widely used Autocorrelation function (ACF) based PDA. The PDA divides the musical signal into frames. For each frame ACF is calculated. The time lag corresponding to highest ACF peak is reported as the estimated pitch period for that frame. The inverse of pitch period is the pitch or fundamental frequency [2]. The algorithm works well on monophonic music.

In Indian classical or film music the main performer or singer is accompanied by other melodic or percussive instruments. Thus the song contains sounds from more than one source, hence classified as polyphonic music. Different instruments may produce sounds which have different pitches or even different frequency partials. Simple PDAs fail in case of polyphonic music. There is a need to develop a PDA which will work accurately in the presence of percussive accompaniment. This paper presents the results of study of the impact of Tabla strokes on pitch estimation of a melodic voice.

The next section discusses the characteristics of various Tabla strokes and classifies them based on acoustic properties. Next the effects of these strokes on pitch estimation using ACF based pitch detector are discussed. We conclude by discussing the usefulness of this study in building a PDA for music with Tabla accompaniment.

2. ACOUSTIC CHARACTERISTICS OF TABLA STROKES

The Tabla is a pair of drums traditionally used for the accompaniment of Indian classical or semi-classical music. The 'bayan' is the metallic bass drum, played by the left hand. The 'dayan' is the wooden treble drum, played by the right hand. A larger variety of sounds is produced on this drum. The drum can be tuned according to the accompanied voice or instrument. It is also important to note that for this class of instruments the main resonance (that corresponds to the main perceived pitch) is much stronger than the resonance of the other harmonics [3]. A mnemonic syllable or bol is associated to each of these strokes. Common bols are : Ge, Ke (bayan bols), Na, Tin, Tun, Tit (dayan bols). Strokes on the 'bayan' and 'dayan' can be combined, like in the bols : Dha (Na + Ge), Dhin (Tin + Ge), Dhun (Tun + Ge).

Strokes Na, Tin, and Tun which are produced on 'dayan' are slowly decaying sounds with many frequency components. Strokes Na and Tin have strong harmonic partials up to 2 kHz while Tun has nearly harmonic partials i.e. the spacing between the successive harmonics is not exactly the same. The difference between these strokes is due to the location of strongest partial among the harmonics. Na has strong partial near 1 kHz, Tun has strong partial near 500

[*] CSE dept.
[†] EE dept.
[‡] KReSIT

Hz while Tin has strong partial below 500 Hz. Stroke Ge produced on 'bayan' has harmonic partials up to 1kHz, and decays slowly. It has very strong partial between 100 Hz to 200 Hz. Strokes Tit produced on 'dayan', and Ke produced on 'bayan' show impulsive nature. They decay very fast. Ke shows predominantly low frequency content while Tit has very weak frequency partials up to 2 kHz. The strokes produced by striking both 'bayan' and 'dayan' simultaneously show characteristics which are mix of separate strokes. For example the spectrum of Dha is similar to the spectrum produced by adding the spectrum of Na and Ge. Most of the Tabla strokes especially the harmonic ones have stronger partials up to 2 kHz. Now we present basis for classification and their conjectured impact on the pitch detection. Details of specific instances of strokes Na, Ge, and Ke are given in Table 1 and Fig. 1 shows their spectrograms.

1. Harmonic structure: The strokes Na, Tin, Tun, and Ge are harmonic and Tit and Ke are inharmonic strokes. The harmonic strokes can confuse pitch detector since simultaneously two harmonic sounds are present, thus multiple pitches exist.
2. Rate of decay: The strokes Ke and Tit are fast decaying and impulsive while strokes Na, Tin, Tun, and Ge are slowly decaying. Fast decaying strokes will affect less number of pitch estimates than slow decaying ones.
3. Location of strong partial: Strokes Ge, Tin, and Tit have strong partials at frequencies below 500 Hz while other strokes have their strong partial above 500 Hz.

3. PITCH DETECTION IN TABLA ACCOMPANIMENT

In order to study the impact of Tabla accompaniment on the pitch estimation of ACF based detector, experiments were performed by mixing Tabla stroke sounds to song and applying ACF based pitch detector. Instead of mixing each stroke three representative strokes Na (Harmonic with high frequency resonance), Ge (Harmonic with low frequency resonance), and Ke (impulsive sound) were chosen for experiments. The Magnitude plots of these strokes are shown in Fig. 1.

3.1. Experiment

Single tabla strokes were isolated in a single channel wave files sampled at 22050Hz, with 16 bits per sample. The acoustic waveforms of these strokes, at rate 1 stroke per second were added digitally in different Signal to Noise Ratio (SNR)s, to acoustic waveform of song 'Ye shaam' sung with syllable /la/ in female voice. The pitch range in the song was 200 Hz to 600 Hz. The total duration of mixed file was 35 seconds. ACF based pitch detector was run on the mixed signal. The frame length was 40 ms with 50% overlapping. Thus pitch was sampled every 20 ms. The pitch estimates were compared with the pitch estimates obtained from pure song file. The term 'pitch error' below means the pitch of mixed signal and pure song didn't match. Table 1 gives the specific details of the Tabla strokes used in experiments.

Stroke	Strongest partial (Hz)	Duration (sec)
Na	790	.45
Ge	137	.35
Ke	-	.1

Table 1. Specific details of Tabla strokes

3.2. Observations

Fig. 3 shows the plots of pitch versus time (upper plot) and plot of power versus time (lower plot) of a short segment of song mixed with strokes Ke, Na, and Ge one every second, and pure song. Pitch errors were observed in the frames in which the ratio of local power of stroke to that of voice is above certain threshold. Thus as the SNR increases the number of errors reduce. The pitch errors can be classified as fine errors and gross errors. Fine errors occur when pitch error is between 3% to 6% of correct pitch, while gross errors occur when that is above 6%. An octave error occurs when estimated pitch is submultiple or multiple of correct pitch. In the case of stroke Na the pitch estimates in first few frames after the onset were mostly found to be near 790Hz which is the strongest partial of Na stroke. Octave errors were observed when the strongest partial (790 Hz) was approximately odd multiple of half the correct pitch estimate. In the case of stroke Ge most of the erroneous pitch estimates were near 137 Hz which was the strongest partial in Ge, while for Ke they were near 172Hz., thus causing gross error. The maximum number of errors were found in case of Na then Ge and least in the case of Ke.

4. DISCUSSION

To understand the behavior of ACF peak-based pitch detection, it is useful to think of the ACF of a signal comprising several frequency components (harmonic and inharmonic), as the inverse Fourier transform of the power spectrum of the signal. The signal power spectrum is insensitive to the relative phases of the components, to the extent that the window is long enough that there is no significant leakage of the frequency components. Due to the linearity of the Fourier transform, the ACF of the signal is the summation of the ACFs of the individual components in the signal power spectrum, and is therefore insensitive to the phase relations between components. For the pure musical tones, each of which contains a number of harmonics, including the funda-

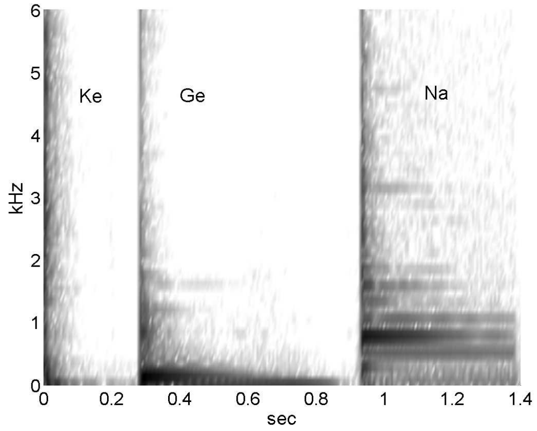
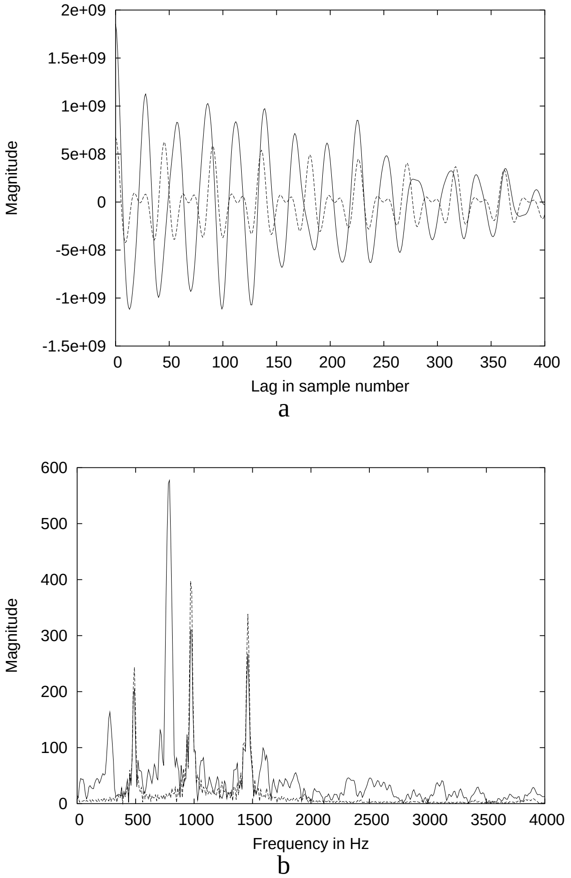


Fig. 1. Spectrograms of Na, Ge and Ke.



a

b

Fig. 2. a. ACF plot and b. Magnitude plot of frame containing the onset of Na with SNR=2. Magnitude plot is displayed for 0-4 kHz frequency range. Correct pitch is 490 Hz (lag 45) and reported pitch 788 Hz (lag 28). Continuous line shows plot of mixed song while dashed line shows that for pure song

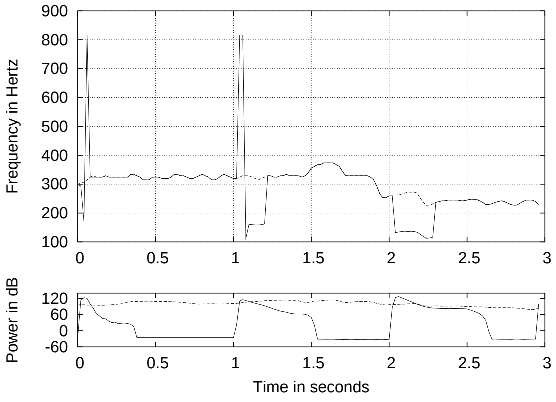


Fig. 3. The dashed line in upper plot shows pitch contour of pure song while continuous line shows that of mixed song. Dashed line in lower plot shows local power of song while continuous line shows that of Tabla stroke. The mixed signal contains strokes Ke for first, Na for second, and Ge for third second

mental, the windowed ACF of the input signal shows peaks at lags corresponding to the pitch period and multiples of the pitch period. The highest peak corresponds to the pitch period, and there is no error in the estimated pitch. On the other hand, when the input signal to the ACF contains noise or interfering partials, there is a perturbation of the peak corresponding to the correct pitch period. The ACF of the interference partial (which can be considered to combine additively with the target ACF to form the corrupted ACF) modifies the values of the original ACF at all lags. Unless the interference partial is very strong, this is not sufficient to change the locations of the prominent peaks (at pitch and pitch multiples) but affects only their relative amplitudes. As a result, the choose the highest peak in the ACF approach typically results in errors due to a misshapen pitch peak or change in the relative amplitudes of peaks leading to pitch octave error [4].

In case of all Tabla strokes one of the partials is very strong compared to other partials. If this partial is strong enough then we can see peaks at lags corresponding to the frequency of this partial only since peaks due to other frequency components are suppressed. Thus the first ACF peak corresponding to strongest partial in Tabla stroke is picked. This can be seen in Fig. 2. This explains the errors occurred at onset of strokes.

The likelihood of octave error increases when the valley of the ACF of strongest partial of Tabla stroke approximately coincides with the first peak of harmonic signal and the peak of strongest partial coincides with the a peak other than first of the signal. This happens when the strongest frequency in stroke is approximately odd multiple of half the correct pitch-frequency. This is shown in Fig. 4. This con-

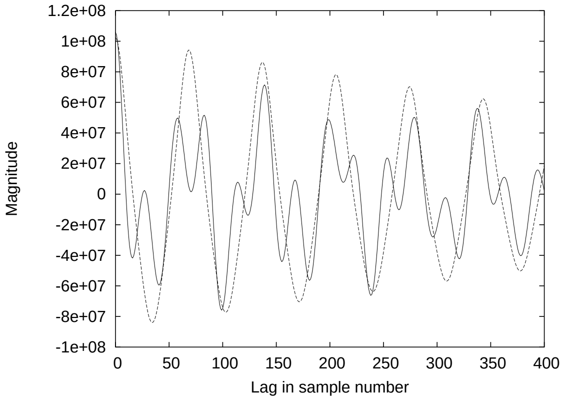


Fig. 4. ACF plot of frame containing stroke Na with SNR=1. Correct pitch is 320 Hz (lag 69) and reported pitch is 158 Hz (lag 139), Continuous line shows plot of mixed song while dashed line shows that for pure song.

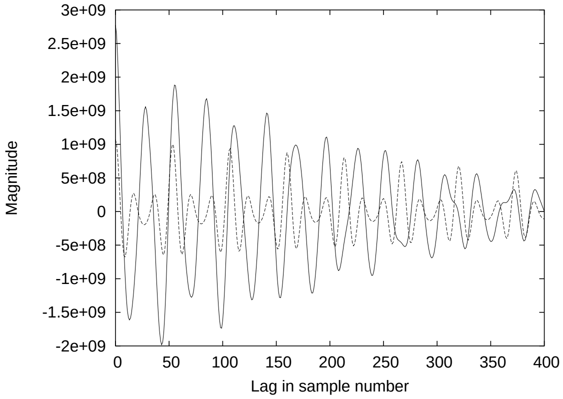


Fig. 5. ACF plot of frame containing stroke Na with SNR=1. Correct pitch is 416 Hz (lag 53) and reported pitch is 393 Hz (lag 56), which is nearly half of 790 Hz. Continuous line shows plot of mixed song while dashed line shows that for pure song.

firms with observations in [4] done with synthetic signals.

When the correct pitch is near some sub multiple of the strongest partial, instead of the strongest partial a pitch near the sub multiple is reported. If the sub multiple and correct pitch are too close then pitch error may be classified as fine error. This can be seen in Fig. 5. This is because the peak other than first (corresponding to the sub multiplication factor) is boosted by the harmonics of fundamental frequency.

It can be easily observed that filtering very low frequency content will minimize or even eliminate the errors in case of strokes with predominant low frequency content. Filtering low frequencies doesn't affect the pitch estimate due to the presence of harmonics at higher frequencies. But just filtering out Tabla sounds is next to impossible since harmonic strokes contain frequencies up to 2 kHz. Also the exact location of the partials varies during the stroke and from instrument to instrument. It can be observed that source of most of the pitch estimation errors are the strongest partials in strokes. This suggests use of some type of magnitude warping or spectral flattening, thus giving more importance to the place of the partial than its strength. We are working in this direction to develop pitch detector.

5. CONCLUSION

This paper presents a brief introduction to the problem of pitch detection in polyphonic music especially with Tabla accompaniment. The importance of pitch detection in AMT and melody retrieval systems was underlined. The impact of Tabla accompaniment on the pitch estimation and its reasons were discussed using ACF based pitch detector. These observations will help us to develop a pitch detector which will work correctly for music with Tabla accompaniment. This paper presents the study of isolated strokes, but it would be interesting to study the complex strokes such as Dhin (Tin + Ge), and words such as TiReKiTa. The strokes may overlap when rhythm is fast. This may have different impact depending upon which strokes overlap, which needs to be studied.

6. REFERENCES

[1] M. A. Raju, B. Sundaram, and Preeti Rao, "Tansen: A query-by-humming based music retrieval system," in *National Conference on Communications, NCC 2003 at IIT Madras*, 2003.

[2] W. Hess, *"Pitch Determination of Speech Signals"*, Springer, New York, 1983.

[3] Gillet O. and Richard G., "Automatic labelling of tabla signals," in *International conference on music information retrieval (ISMIR)*, 2003.

[4] Preeti Rao and Saurabh Shandilya, "On the detection of melodic pitch in a percussive background," *J. Audio Eng. Soc.*, *Vol. 52*, 2004.