Pitch tracking of voice in tabla background by the two-way mismatch method

Ashutosh Bapat[*], *Preeti Rao*[†]

Indian Institute of Technology Bombay, India

(email: prao@ee.iitb.ac.in)

Abstract

Obtaining the detailed pitch contour of the melody from audio recordings of Indian classical music is important both from a pedagogical as well as musicological perspective. In this work, the problem of pitch tracking of the singing voice in percussive accompaniment is considered. 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. The acoustic signal characteristics of the percussive accompaniment that pose specific challenges to conventional pitch detection algorithms (PDAs) are discussed. An experimental investigation of the performance of a frequency-domain PDA, the two-way mismatch method, is carried out for a variety of simulated and real music signals of singing voice in tabla accompaniment. A post-processing method based on dynamic-programming based smoothing is proposed and shown to significantly improve the accuracy of the estimated pitch contour.

Keywords: Pitch tracking, dynamic programming, Indian classical music, two-way mismatch method

1. Introduction

Music transcription refers to the conversion of an acoustic musical signal to a symbolic representation. It is an important exercise both in the pedagogy of music as well as for musicological studies. While the human auditory system remains the most reliable transcriber of music, there has been much recent research on automatic music transcription systems. These systems combine digital signal processing of the acoustic signal with pattern classification techniques to come up with musically meaningful symbolic representations. However despite many efforts and some successes, no practical general-purpose system exists that can accurately transcribe a wide variety of music. Indian classical music poses its own peculiar challenges to the automatic music transcription task. While the lack of an accepted symbolic notation (like in Western music) is a significant obstacle, almost equally challenging is the signal analysis problem of extracting the musically relevant parameters.

Acoustic signal analyses of music can yield information on the various aspects of melody, rhythm as well as the timbre of the instruments playing. Of these, the melody is arguably the single most important descriptor of a piece of music. The melody is represented by the variation of the perceived pitch (often assumed to be the fundamental frequency of the periodic signal) with time [1]. Pitch detection is the process of estimating an instantaneous pitch value from the signal at discrete intervals of time. The output of a pitch detector is typically a pitch contour i.e. a trace of the pitch as it evolves with time.

Several pitch detection algorithms (PDAs) have been proposed over the last three decades [1]. These have been broadly classified as being based either on measuring time-domain periodicity or on harmonic pattern matching. The autocorrelation function (ACF) PDA is a well-known example of short-term periodicity measurement via time domain correlation. Sub-harmonic summation (SHS) represents a frequency domain PDA based on harmonic structure detection. While most PDAs provide accurate estimates with monophonic (single voice) signals, the reliable pitch estimation of polyphonic music remains a challenging problem [2, 3].

In Indian classical music, typically, a single melodic instrument is accompanied by a percussive instrument such as the tabla providing the rhythm. Also present throughout the performance is the tanpura (drone) sounding the selected tonic note. Thus a typical audio recording contains sounds from more than one source, and can therefore be considered polyphonic. Polyphonic music such as this presents a particular situation of pitch tracking in noise or interference, where the interference is from the accompanying instruments. That a reliable pitch detection method would indeed be very valuable is evident from the observation that musicological studies of Indian classical music are often unable to exploit the rich sources of commercial audio recordings due to the presence of instrumental accompaniments [4].

The present work addresses the problem of reliable pitch tracking of a single melodic instrument, more specifically,

[*] Department of Computer Science and Engineering

[†] Department of Electrical Engineering

the human voice, in the presence of tabla accompaniment. The two-way mismatch (TWM) method [5], a frequency-domain PDA known to perform well on monophonic music signals with realistic degradations is investigated for this application. A new post-processing method based on dynamic programming with smoothness constraints is proposed to improve the accuracy of TWM pitch estimates in the presence of strong interference. An experimental evaluation is presented of the complete pitch tracker on a data set of realistic signals.

2. Signal characteristics

In the present context, the melody is solely carried by the singing voice. Hence “target” is the singing voice. The “interference” is the tabla strokes.

2.1. Characteristics of singing voice

In Indian classical music, the pitch contour of singing voice is not made up of discrete horizontal lines corresponding to distinct and steady notes in the melody but it is a continuously evolving curve. The pitch contour shows many types of pitch inflexions like glides, oscillations, bends, stresses, and accents, all collectively called *gamakas*. They are important from the perspective of the emotions song conveys and also in Indian classical music they are important for appreciation of *ragas*. A detailed study of *gamakas* and their importance in various *ragas* is presented in [4, 6]. The work presented in this paper is focused on the vocals which contain only vowels or semivowels such that every sung segment has a clear pitch contour.

2.2. Acoustic characteristics of tabla strokes

The tabla is a pair of drums traditionally used as rhythmic accompaniment in 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 and produces a large variety of sounds. The drum can be tuned according to the accompanied voice or instrument. It is important to note that for this class of instruments the main resonance is much stronger than the resonances of other harmonics [7]. 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).

The tabla sounds typically have low pass spectra overlapping that of human voice. Strokes Na, Tin, and Tun which are produced on ‘dayan’ are slowly decaying sounds with many frequency components. They have many harmonic partials, with a single partial stronger than others. The difference among these strokes is due to the location of the strongest partial among the harmonics. Na has its strongest partial near 800 Hz, Tun has its strongest partial near 500 Hz, while Tin has its strongest partial below 500 Hz. Stroke Ge produced on ‘bayan’ has harmonic partials up to 1 kHz, and decays slowly. It has very strong partial between 100 Hz to 200 Hz. In case of Ge the harmonics other than the strongest one decay relatively fast. Strokes Tit produced on ‘dayan’ and Ke produced on ‘bayan’ show impulsive nature. They decay very fast. Ke shows a noisy spectrum with no clear partials 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 a mix of strokes produced on both drums separately. For example, the spectrum of Dha is similar to the spectrum produced by adding the spectra of Na and Ge. Most of the tabla strokes especially the harmonic ones have stronger partials up to 2 kHz. To summarise above discussion, tabla strokes can be classified on the basis given below.

1. Harmonic structure: The strokes Na, Tin, Tun, and Ge are harmonic and Tit and Ke are inharmonic strokes.

2. Rate of decay: The strokes Ke and Tit are fast decaying and impulsive while strokes Na, Tin, Tun, and Ge are slowly decaying.

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.

Based on this, experiments presented here use three representative strokes Na, Ge, and Ke which cover all the characteristics of tabla strokes. Details of specific instances of strokes Na, Ge, and Ke are given in Table 1 and Fig. 1 shows their spectrograms.

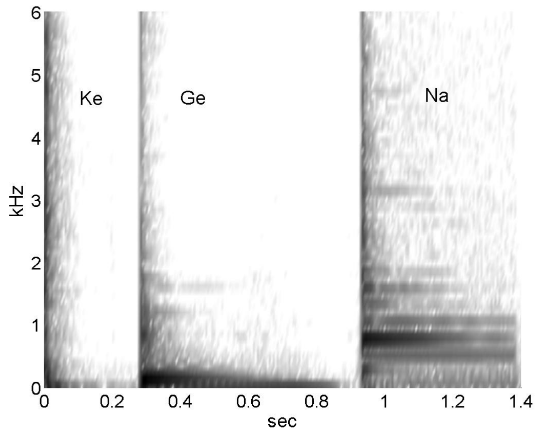


Fig. 1. Spectrograms of tabla strokes Na, Ge and Ke

Fast decaying strokes will affect less number of pitch estimates than slowly decaying ones. The harmonic strokes

Stroke	Strongest partial (Hz)	Duration (sec)
Na	790	.45
Ge	137	.6
Ke	-	.2

Table 1. Specific details of tabla strokes

can confuse pitch detector since simultaneously two harmonic sounds are present, thus multiple pitches exist.

Fig. 2 shows spectrogram of a portion of song in female voice mixed with tabla stroke Na. The continuous horizontal curves denote the spectral evolution of voice. The vertical strikes at 6 and 7 second indicate the onset of stroke Na. We also observe strong partials of Na at 523 Hz and 790 Hz interspersed with the voice harmonics.

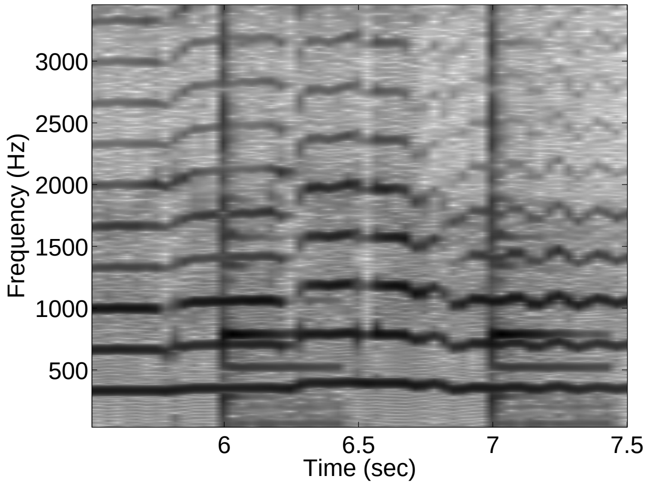


Fig. 2. Spectrogram of four notes sung by female singer mixed with stroke Na.

3. Two-way mismatch procedure

In the two-way mismatch procedure the pitch of a signal is estimated by choosing the fundamental frequency which minimizes the discrepancy between the measured spectrum and the spectrum predicted for that fundamental frequency. The musical signal is divided into 40 ms long frames with 50% overlap. The spectral peaks, i.e local maxima above a certain threshold value relative to the global maximum in the magnitude spectrum of each frame are stored. These peaks are henceforth called “measured” peaks.

The procedure employs two mismatch error calculations as depicted in Fig. 3, one based on the difference between each measured partial and its nearest predicted partial, and other based on the difference between each predicted partial and its nearest measured partial. This two way mismatch helps to avoid the octave errors by applying a penalty for partials those are present in the measured data but not in predicted data (case with multiple type octave error), and also for the partials whose presence is predicted but do not appear in measured data (case with sub-multiple type octave error).

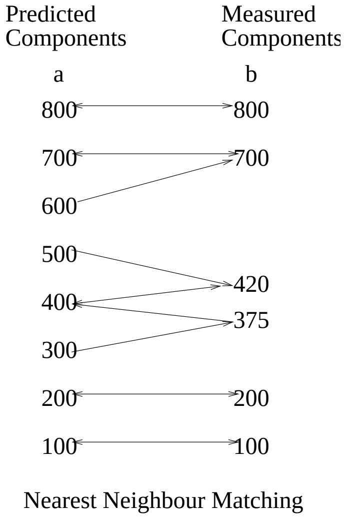


Fig. 3. Two step TWM error calculation at assumed trial fundamental frequency of 100 Hz. 1. Each measured partial is compared with its nearest predicted neighbour, indicated by arrows from b to a, 2. Each predicted partial is compared with its nearest measured partial, indicated by arrows from a to b. The total error is normalized sum of these errors.

The two errors are computed as follows

- Predicted to measured error

$$ND(n) = \Delta f_n^p . (f_n^p)^{-p}$$

$$E_{p \rightarrow m} = \sum_{n=1}^{N} ND(n) + (\frac{a_n}{A_{max}}) * (q.ND(n) - r) \quad (1)$$

Where f_n^p is the frequency of n^{th} predicted partial for some trial fundamental, Δf_n^p is the difference between the frequency of predicted partial and its closest measured partial. a_n is the amplitude of the closest measured partial to f_n^p. A_{max} is the maximum of the amplitudes of measured partials. N is smallest integer greater than f_{max}/f_0 , where f_{max} is maximum of the frequencies of measured patials and f_0 is trial fundamental frequency.

- Measured to predicted error

$$ND(k) = \Delta f_k^m.(f_k^m)^{-p}$$

$$E_{m\rightarrow p} = \sum_{k=1}^{K} ND(k) + (\frac{A_k^m}{A_{max}}) * (q.ND(k) - r) \quad (2)$$

Where f_k^m and A_k^m are the frequency and amplitude of k^{th} mesured partial, Δf_k^m is the difference between the frequency of measured partial and its closest predicted partial. A_{max} is the maximum of the amplitudes of measured partials. K is number of measured partial.

Total TWM error is given by,

$$E_{total} = \frac{E_{p\rightarrow m}}{N} + \rho\frac{E_{m\rightarrow p}}{K} \quad (3)$$

Values of parameters q, r, ρ are kept same as given in [5]. The paper suggests a value of parameter $p = 0.1$ to reduce emphasis placed on the low level, high frequency fundamentals. Same value of p is used here.

The error is calculated for a series of trial fundamental frequencies spanning the known range of pitches in input signal. The spacing between the trial frequencies can be chosen to achieve required accuracy for estimates. Finally for our pitch tracker the E_{total} for each trial frequency is normalised by the maximum error in that frame. The normalised error is henceforth called TWM error for simplicity.

The plot of TWM error versus trial fundamentals for a complex tone of 300 Hz is shown in Fig. 4. The plot shows a very deep minimum at 300 Hz. Also the minima at other trial frequencies like 450 Hz and 600 Hz, which are factors of partial frequencies in the complex tone, can be seen. Out of numerous local minima in the plot, the trial frequency corresponding to global minimum is chosen as the pitch for the frame. Fig. 4 shows similar plot for a frame of signal obtained by same complex tone is mixed with tabla stroke Na. It can be observed that the minimum at 300 Hz has become shallower significantly and there is a minimum at 261 Hz which is the fundamental frequency of Na. The global minimum at 100 Hz is deeper than both of these minima.

Similar to spectrogram, TWM error for trial frequencies at various instances of time is plotted. Such a plot is shown in Fig. 5 for a complex tone of 300 Hz mixed with tabla stroke Na. This correlogram shows a very dark harizontal line at 300 Hz where only pure complex tone is present, i.e. from 0 to .2 seconds and from .65 to 1 seconds. But wherever stroke Na is present, i.e. from .2 to .65 seconds, this line has become brighter, indicating that the normalised error has increased at 300 Hz in those frames. At the onset of Na this dark line has become much brighter for a couple of frames. It can be observed that low frequency region is darker where the stroke is present than where it's absent.

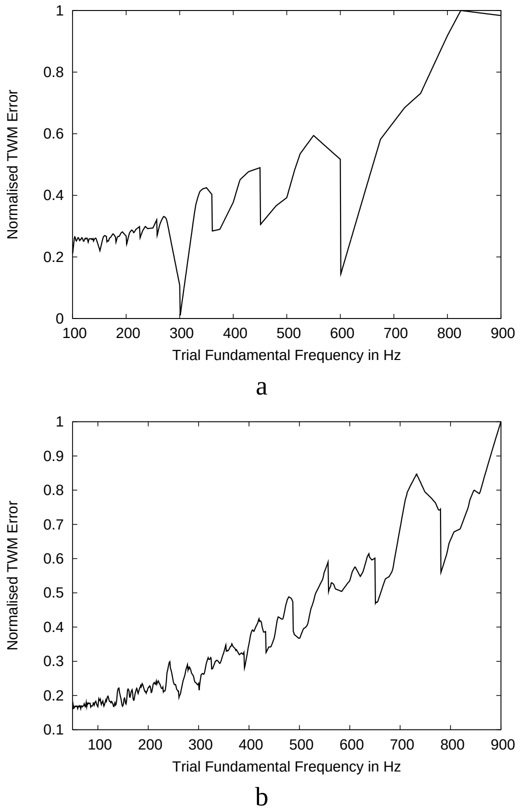


Fig. 4. a. TWM Error plot for complex tone of 300 Hz containing harmonics at 600, 900, 1200, 1500, 1800 Hz all having equal amplitude. b. Similar plot for the same complex tone mixed with tabla stroke Na.

The pitch contour of a song with tabla strokes added to it, as obtained from the TWM PDA, is shown in Fig. 6.a. The pitch contour of the mixed song shows sharp variations in pitch at the onset of tabla stroke as well as during the tabla stroke. Pitch in pure song does not vary so drastically in successive frames. The values of pitch in adjacent frames are expected to be strongly correlated [8]. To rectify these errors postprocessing as explained next, is applied.

4. Postprocessing

Postprocessing based on dyanamic programming is very common in a typical pitch tracking system [1, 8, 9]. Though postprocessing is generally used for overcoming anomalies of vocal cords or irregularities in signal, here it specifically designed to overcome the significant impact of tabla strokes.

The block diagram of the pitch tracker is shown in Fig. 7. Postprocessing involves DP based smoothing and pitch correction. Postprocessing is applied to segments of songs which contain singing voice. For each continuous segment of singing voice a we calculate two dimensional error

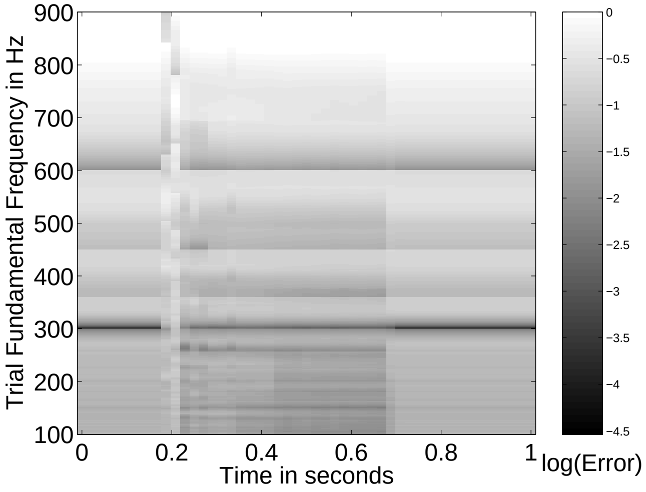


Fig. 5. Correlogram of the log of normalized TWM error of a complex tone of 300 Hz as described in Fig. 4.a, mixed with tabla stroke Na. The stroke starts about .2s and lasts for .45s. The darkness of a pixel $d(p, t)$ indicates the log of normalised error for frequency p at time instance t, according to the scale shown on the right. Thus darker the point lower is error.

matrix E, where a cell $E(p, j)$ gives the TWM error for trial frequency p for frame j. Thus j^{th} column of matrix gives TWM error values for various trial frequencies for j^{th} frame. In this section, we discuss DP based approach to recover a smooth (and possibly more accurate) pitch contour by eliminating the erratic variations due to the presence of tabla interference.

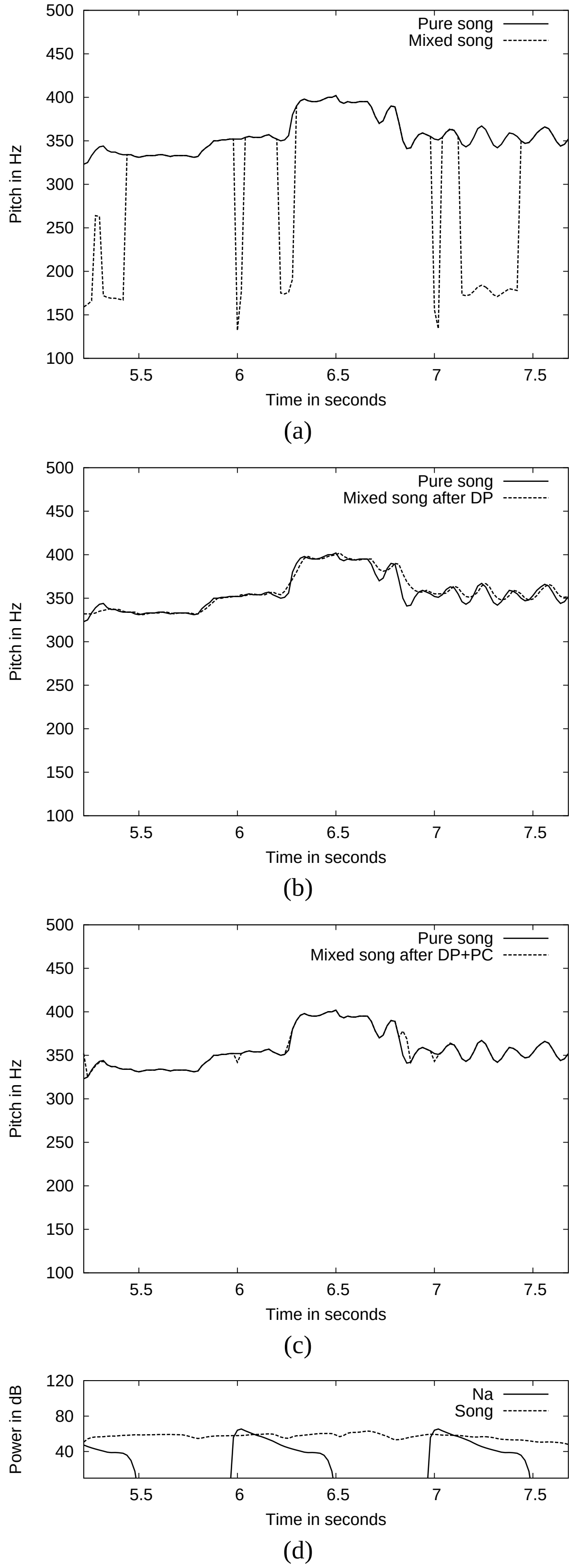
4.1. Dynamic programming based smoothing

DP approach used here is similar to that used in [8]. It consists of three essential parts. First, a measurement cost for the estimated pitch. Then, a smoothness cost for time evolution of pitch. These two costs make up the local transition cost. Finally, an optimality criterion to represent the trade off between the measurement and smoothness cost is defined in the form of global transition cost. The DP strategy is used to carry out optimization of the global transition cost. The measurement cost is provided by the error matrix E. The smoothness cost assumed by us is given by

$$W(p, p') = 1 - \frac{e^{\frac{-(p'-p)^2}{2\sigma}}}{s} \qquad (4)$$
$$\sigma = c * p$$

where p and p' are pitch estimates in successive frames in that order. The local transition cost is thus given by

$$c(p(j), p(j-1), j) = E(p(j), j) + W(p(j-1), p(j)) \quad (5)$$



(a)

(b)

(c)

(d)

Fig. 6. a. Pitch Contours of portion of song (pure and mixed with stroke Na) obtained by TWM based PDA. b. Pitch Contour of same sample after applying DP. c. Pitch Contour of same song after applying pitch correction. d. Power contour of stroke Na and the song.

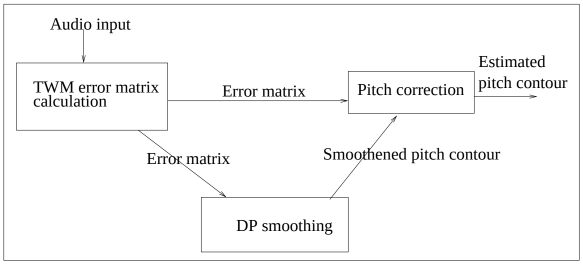


Fig. 7. Blocks for Pitch Tracker

with $c(p(1), p(0), 1) = E(p(1), 1)$. Here $p(j)$ is the pitch estimate at frame j. The global transition cost is given by,

$$S(p(1), .., p(j), .., p(N)) = \sum_{j=1}^{N} c(p(j), p(j-1), j) \quad (6)$$

, where N is the number of frames in the given continuous segment of song. The pitch contour with least global transition cost among many possible pitch contours for a given segment of song is chosen as the pitch contour estimate of that segment.

The smoothness cost function is obtained by modifying the equation of a Gaussian distribution with σ being standard deviation and p being mean. σ is proportional to the pitch estimate in current frame through constant c. This is in keeping with the fact that variation in the pitch is expected to be more at higher pitch values than that at lower pitch values. It can be noted that the smoothness cost for a transition from higher pitch (say 300 Hz) to lower pitch (say 50 Hz) is lower than a transition from lower pitch to higher pitch. This may cause DP smoothing to get trapped into lower pitch values. Hence smoothness cost function which saturates for larger pitch transitions, is used. Constant s controls the minimum value of W. The values $s = 1.3$ and $c = .9$ are experimentally found to give good results. An example of the W function is shown in Fig. 8 for $p = 100$.

The DP algorithm can be understood by considering a state space, $\{(p, j)|p \in \{trialfundamentals\}, j = 1\, to\, T\}$ as shown in Fig. 9. The pitch contour can be seen as the path ((p(1),1), (p(2), 2), ..., (p(j), j), ...(p(T), T)) through this state space, where $p(j)$ is pitch estimate at j^{th} frame. While passing through each state (p, j) path incurs a cost of $E(p, j)$ and while making a transition from (p, j) to $(p', j + 1)$ it incurs a cost of $W(p, p')$. Only transitions allowed are from (p, j) to $(p' j{+}1)\, where\, p, p' \in \{trialfundamentals\}$. Thus global transition cost is the cost of a path passing through this state space. DP formulation is used to find the minimum cost path through the state space.

Effect of applying DP on the pitch contour can be seen in Fig. 6.b. It can be observed that DP gives a smoothened pitch contour, but at the cost of suppressing certain fast variations in the pitch. Lower the value of c higher is the smoothning of the pitch contour.

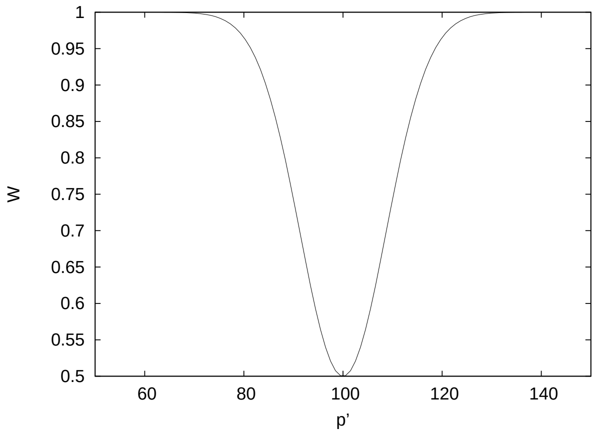


Fig. 8. Shape of smoothness cost function W, $for\ s = 2, \sigma = 75$

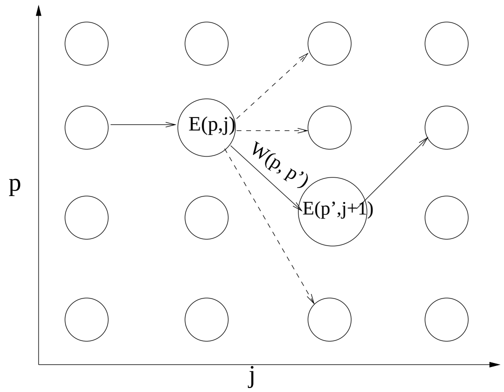


Fig. 9. State space representation of dynamic programming. The states and edges are labeled by their costs. The dashed edges indicate the possible transitions for state (p, j) while the solid edges indicate the minimum cost path found by dynamic programming.

4.2. Pitch Correction

Pitch correction makes use of the fact that TWM error function contains a local minimum near the correct pitch even if the frame is corrupted by interference. The deepest local minimum in a certain range (within $\pm 6\%$ of the estimated pitch) near the pitch estimated by DP is searched. The frequency corresponding to this minimum is declared as the correct pitch. In case, no local minimum exists in the search range, the pitch estimated by DP is kept unchaged. Pitch contour obtained after applying Pitch Correction on estimates found by DP is shown in Fig. 6.c.

4.3. Effects of DP and pitch correction

When the pitch changes rapidly, the pitch contour estimated by DP 'lags behind' the correct pitch contour. This can be seen in Fig. 6.b. This can be observed even in those portions where tabla stroke is absent. This phenomenon occurs when the smoothness cost dominates the total cost at the correct pitch, due to the large difference in value from the pitch of the previous frame. The effect accumulates causing a lag in the pitch contour estimated by the DP. If the pitch contour is too steep, DP smoothing may completely loose track of the correct pitch contour. Sometimes the correct pitch contour may show a sharp peak (or valley) as shown in Fig. 6.b. A peak and adjacent valley can be observed about 6.75 seconds in this plot. When DP is applied on such a portion of song, the cost of the correct pitch contour is higher than that of the straight path (or "short cut") taken by DP, because of the total smoothness cost incurred due to the sharp changes in the pitch. Thus application of DP smoothing may eliminate sharp peaks or valleys, introducing errors in the pitch estimation.

Pitch correction eliminates some of the errors caused by DP smoothing. More the the search range of pitch correction, more errors are corrected. However if the search range is too wide, pitch correction may pick up a deeper minimum caused due to interference in frames corrupted by tabla strokes. It can be seen in Fig. 6.c that pitch correction has corrected most of the errors but is unable to correct some of the errors in portion around 6.75 seconds.

5. Experimental evaluation

The pitch tracker has been tested on sample data prepared by mixing singing voice waveforms with waveforms of tabla strokes. The songs were sung in syllable /la/ and /aa/. Three different tabla strokes Na, Ge, and Ke were digitally added separately to these song waveforms in global SNR 2, 2, and 3 and at rate of 1, 1 and 2 per second equally spaced respectively. Since tabla stroke Ke is very short in duration as compared to other two strokes, it was mixed with higher rate, so that roughly half the frames of song remain unaffected. For the same reason Ke was added at higher global SNR. The songs and tabla strokes were sampled at 22050 Hz using 16 bits per sample. The pitch range in the songs was 200 to 600 Hz. Frame length for all algorithms was 40 ms, with 50% overlap and pitch search range given as 100 Hz to 900 Hz. The songs have been selected so as to cover most of the types of variations of pitch as mentioned in Sec. 2.1. It is assured that the tabla strokes overlap various types of pitch variations.

The following algorithms were applied on the test samples.

TWM: Two way mismatch algorithm with our parameter values.

TDP: TWM followed by DP as explained in Sec. 4.1.

TDC: TWM followed by DP and pitch correction as explained in Sec. 4.2.

It is assumed that the pitch estimates obtained from pure songs by applying TWM method are correct. An error is said to occur when there is difference between pitch estimates obtained from mixed sample and those from its corresponding pure sample. Since error within 3% of correct pitch estimate is considered to be within half semitone, it does not cause a note to change. The errors within 3% to 6% of the correct pitch estimates are called 'fine errors'. The errors above 6% of the correct pitch estimates are called 'gross errors'. The error rates have been obtained for ten song samples of average length 20s. Table 2 gives error rates in pitch estimates by the algorithms discussed above, of two representative songs added with three strokes separately. Errors are counted only in the frames affected by the tabla strokes.

	TWM		TDP		TDC	
	Fine	Gross	Fine	Gross	Fine	Gross
Na	0.0	49.3	4.6	13.4	2.1	14.8
Ke	0.0	20.9	3.9	2.1	3.4	2.5
Ge	0.0	25.7	4.9	5.1	0.2	5.1

(a.) A song sung in /la/,
479 frames corrupted by Na, 669 by Ke, 587 by Ge

	TWM		TDP		TDC	
	Fine	Gross	Fine	Gross	Fine	Gross
Na	4.7	14.7	11.6	1.6	5.3	2.1
Ke	0.0	22.5	8.1	4.4	1.5	4.1
Ge	0.0	17.9	7.2	2.5	0.0	2.5

(b.) A song sung in /aa/,
190 frames corrupted by Na, 271 by Ke, 235 by Ge

Table 2. Error rates (in percentage) for two representative songs. a. song with many fast transitions in pitch, b. song with slowly varying pitch contour

6. Results and discussion

It can be seen from Table 2 that, there is improvement in the error rates in pitch estimation by TDC and TDP over the TWM algorithm. Error rates in Table 2 are representative of ten selected song samples. It can be observed that the application of DP smoothing has brought down the number of gross errors significantly, but it has increased the number of fine errors. Application of pitch correction further has brought down the number of fine errors.

We note that, sometimes application of pitch correction causes an increase in the number of gross errors. In such cases the TWM error function of the corrupted frame shows a minimum in the pitch correction search range that is deeper than that at the correct pitch. In this case the pitch correction chooses this local minimum causing a gross error.

Errors that remain uncorrected after applying TDC or TDP are found to be located in the segments of song where there is sharp change in pitch. A particular example of such a case is shown in Fig. 10. It can be seen that, in the pitch contour of pure song that the pitch has changed by 150 Hz in about .3 seconds in portion 19.1 to 19.4 seconds, which is very steep change. TDP algorithm has completely eliminated a small step in this portion with a pitch about 300 Hz. The error introduced by DP in this portion is too large to be corrected by the pitch correction module.

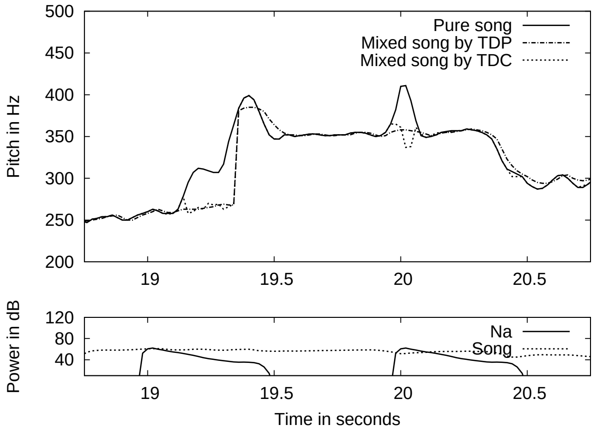


Fig. 10. Pitch contours of a portion of song mixed with tabla stroke Na obtained by TDC and TDP. The solid line indicates the pitch contour of pure song obtained by TWM. Lower plot gives power contour of stroke Na and the song

As discussed in Sec. 4.3, application of DP may also introduce errors in the portions of the song where tabla stroke is absent. This can be corrected by choosing pitch estimates obtained from TWM PDA for frames not containing tabla strokes, and choosing pitch estimates obtained by DP and pitch correction for frames affected by tabla strokes. This requires a strategy to be developed for classifying frames according to the presence or absence of tabla.

7. Conclusion

The problem of pitch tracking of the singing voice in typical Indian classical music recordings is considered. The acoustic signal characteristics of the percussive accompaniment that pose specific challenges to conventional PDAs are discussed. An experimental investigation of the performance of a frequency-domain PDA, the Two Way Mismatch method, is carried out for a variety of simulated and real music signals of singing voice in tabla accompaniment. The spectral structure of tabla strokes causes intermittent pitch estimation errors. A dynamic-programming based smoothing algorithm is proposed to improve the reliability of the pitch contour while retaining rapid variations in pitch as far as possible. This post-processing of the TWM-estimated pitch contour is shown to achieve significant reduction in the pitch error rates. Further, the proposed pitch-tracking system offers a number of parameters which can be tuned for improved performance in specific signal settings.

8. References

[1] W. Hess, “Pitch determination of acoustic signals - an old problem and new challenges,” in *Proc. ICA (International Congress on Acoustics), Kyoto, Japan*, 2004.

[2] A. Klapuri, T. Virtanen., and J. Holm, “Robust multipitch estimation for the analysis and manipulation of polyphonic musical signals,” in *Proc. DAFX-00*, 2000.

[3] P. Rao and S. Shandilya, “On the detection of melodic pitch in a percussive background,” *J. Audio Eng. Soc., Vol. 52*, 2004.

[4] M. Subramanian, “An analysis of gamakams using the computer,” *Sangeet Natak, Vol.XXXVII, pp 26-47*, November, 2002.

[5] R. Maher and J. Beuchamp, “Fundamental frequency estimation of musical signals using a two-way mismatch procedure,” *J. Acoust. Soc. Am. Vol 95(4)*, pp. 2254–2263, April, 1994.

[6] A. Krishanswamy, “Pitch measurement verses perception of south indian classical music,” in *Proc. Stockholm Music Acoustic Conference (SMAC03)*, August 6-9, 2003.

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

[8] H. Ney, “Dynamic programming algorithm for optimal estimation of speech parameter contours,” *IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-13, No. 3, pp. 208-214*, March-April 1983.

[9] B. Secrest. and G. Doddington, “Postprocessing techniques for voice pitch trackers,” in *Proc. International Conference on Acoustics Speech and Signal Processing (ICASSP82), pp 172-175*, 1982.