Proc. of Music-AI (International Workshop on Artificial Intelligence and Music) in IJCAI, 2007, Hyderabad, India.

Melodic Contour Extraction for Indian Classical Vocal Music

Ashutosh Bapat[1], Vishweshwara Rao[2], and Preeti Rao[2]

[1] CSE Department, Indian Institute of Technology-Bombay, India
[2] EE Department, Indian Institute of Technology-Bombay, India
{vishu, prao}@ee.iitb.ac.in

Abstract. The problem of pitch tracking of the singing voice in the presence of Indian percussive interference, specifically the tabla, is considered. To overcome the problems due to this particular type of interference, a pitch tracker is used that applies dynamic programming (DP) based smoothing on pitch estimates obtained from a spectral-domain pitch detection algorithm (PDA) that uses harmonic matching. Experiments on real and simulated signals show the superiority of the spectral domain PDA over a correlation domain PDA in terms of pitch detection accuracy and suitability of the PDA output for post-processing. A new smoothing cost function is proposed and evaluated. The paper formulates general rules guiding the choice of cost functions participating in the DP based post-processing for this particular problem.

Keywords: Pitch tracking, Indian music, Dynamic programming

1 Introduction

Obtaining accurate, high resolution pitch tracks of the melodic instrument is an important pre-requisite for musicological studies of Indian classical instrumental or vocal music [1] [2] [3]. While a number of automatic pitch detection methods are available for speech and music applications, what makes this task particularly challenging is the typical Indian classical music setting where the voice (in vocal music) is accompanied by a drone such as the *tanpura*, providing the fixed tonic, and rhythm provided by a percussive instrument such as the *tabla*. The presence of accompaniment leads to errors in the detected pitch of common pitch detectors, which otherwise perform well with purely monophonic signals.

The tanpura, an overtone-rich stringed instrument, is tuned to the tonic chosen by the singer and is heard as a constant background throughout the performance. Its prominence in terms of its audibility relative to the voice is more due to its timbre rather than the strengths of its partials. It is found that in segments where only the tanpura coexists with the singing voice, the vocal pitch is accurately tracked by common pitch detection algorithms (PDAs). However, the presence of percussion, in the form of the intermittent tabla strokes, causes significant noise-like degradation of the pitch estimates. PDAs generally employ post-processing of the local pitch estimates based on an assumed smoothness of the pitch contour. For instance, median filtering can easily correct for isolated pitch jumps while lowpass filtering corrects corrupted values by interpolating from temporally adjacent regions. However such techniques are not as effective in the Indian classical music setting due to the possibility that important pitch variations may be completely obscured by the longer duration tabla strokes. Unlike Western music, which is grounded on the tempered scale with clear, distinct intervals separating notes, Indian classical music is characterized by the dominating presence of pitch glides and inflections.[1] The pitch inflections, clearly perceived and recognized by experienced listeners, serve an important aesthetic function within the melodic contour and therefore need to be captured accurately [2] [3]. A more systematic approach to post-processing are dynamic-programming (DP) based methods which take into account candidate pitch estimates from the PDA other than just the locally optimal estimate. This amounts to combining suitably defined local measurement and smoothness costs into a global cost, which can then be optimized over a continuous voiced segment by the use of DP [4].

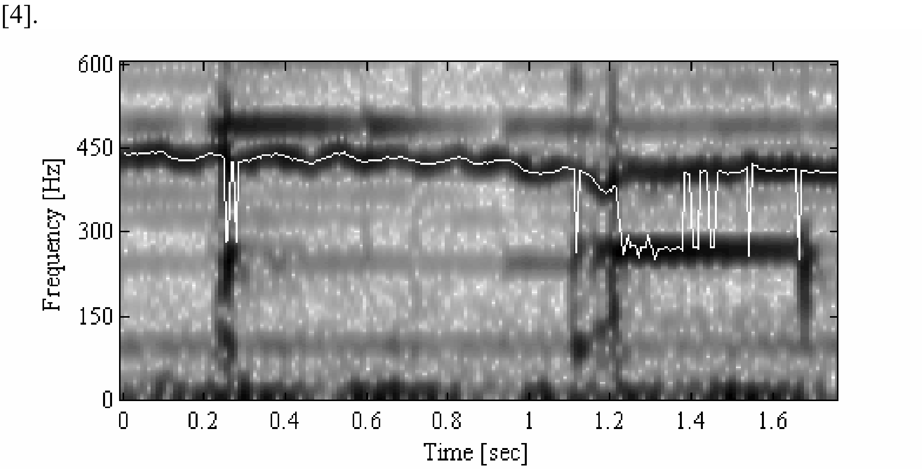

Fig. 1. Pitch contour *(white line)* as detected by a modified ACF PDA superimposed on the zoomed in spectrogram, of a segment of Indian classical music that contains a female voice and a drone throughout and Tabla strokes in some regions.

In the next section, the problem that the percussion poses is illustrated by an example and the salient characteristics of a common percussive stroke, found to cause persistent errors in the estimated pitch contour, are reviewed. We next investigate the performance of a harmonic matching based PDA on a set of simulated signals, and compare it with the more common autocorrelation function (ACF) based PDA. Sec. 4 considers a DP-based post-processing technique, and compares the suitability of the PDAs in this context. A new smoothness cost function is derived from observations on a training data set of classical singing, and evaluation results are reported. The paper concludes by formulating general rules guiding the choice of cost functions for DP-based post-processing.

[1] Related audio examples, figures and experimental results are available at http://www.ee.iitb.ac.in/uma/~daplab/PitchTracking/index.htm

2 Percussion Signal Characteristics

Typical Indian percussion (here, tabla) consists of a wide variety of strokes, each labeled with a mnemonic. Two broad classes are: 1. tonal strokes that decay slowly and have a near-harmonic spectral structure (thus eliciting a pitch percept) and 2. impulsive strokes that decay rapidly and have a noisy spectral structure. The spectra of some tonal strokes, soon after the onset, exhibit harmonics that lie in the same frequency range as those of the singing voice. During onset, the local Signal-to-Noise Ratio (SNR) can dip as low as -5 dB. Subtractive schemes to suppress the percussion prior to pitch detection are not expected to be effective due to the variety of tabla strokes and the high acoustic variability of any given stroke [5].

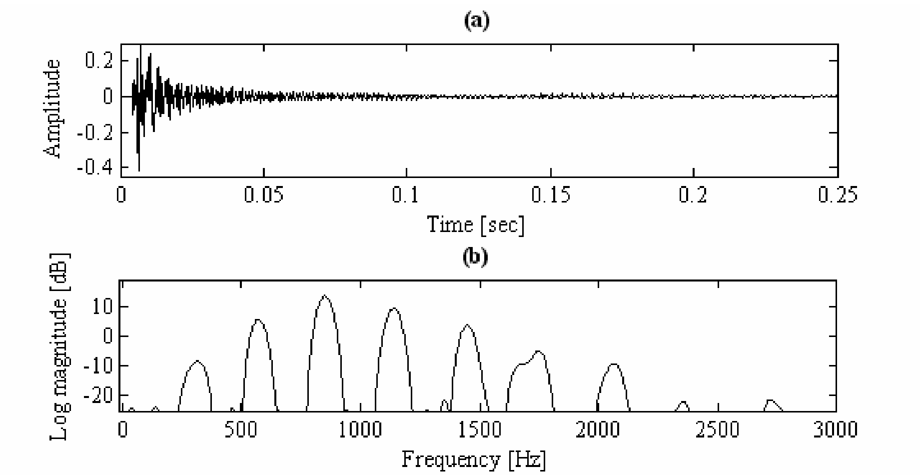

Fig. 2.a. Time domain waveform of a typical Na stroke **b**. Spectrum of a typical Na stroke immediately after its onset. A near-harmonic structure is clearly visible.

An illustration of the degradation caused by tabla percussion is seen in Fig. 1, showing the pitch contour estimated by a modified ACF PDA [6], with recommended parameter settings. The estimated pitch contour is superimposed on a spectrogram of the signal, a typical classical vocal recording segment. In this segment, the sequence of tabla strokes is as follows: impulsive stroke (0.22 sec), impulsive stroke (1.15 sec), tonal stroke (1.2-1.7 sec), and impulsive stroke (1.7 sec). The impulsive strokes appear as vertical narrow dark bands. The tonal stroke (associated with the mnemonic 'Tun') is marked by the presence of a dark (high intensity) horizontal band around 290 Hz, which corresponds to its fundamental frequency. The other, relatively weak horizontal bands correspond to tanpura partials. We note that all the strokes degrade the performance of the PDA, which is otherwise able to accurately track the pitch of the voice in the presence of the tanpura, as indicated by the region where the pitch contour overlaps with the dark band in the spectrogram corresponding to the voice fundamental frequency (between 0.4 and 1 seconds). While the errors due to the impulsive strokes are localized, the tonal stroke causes errors that are spread over a long segment of the pitch track. These latter errors are not just voice octave errors but interference errors i.e. when the pitch estimated is actually the interference pitch, indicated by the lower dark band present temporarily between 1.2 and 1.7 seconds. A tonal stroke is selected for further study because this class of strokes cause extended duration errors which cannot be corrected by common smoothing techniques.

The acoustic characteristics of various tabla strokes were studied from [5]. In particular, the stroke associated with the mnemonic ‘Na’ is chosen because (a) it is commonly occurring, (b) it is slowly decaying (Fig. 2a), lasting for maximum durations of about 0.75 sec and (c) immediately after its onset, its spectrum shows the largest number of significant partials in a near-harmonic relationship (Fig. 2b) It is observed that just around the onset, partials in the frequency ratio 2:3:4:5 are the most significant, and that the second partial is always at least 5 dB stronger than the rest. Generally, these partials remain the most significant throughout the duration of the sound produced by the stroke. Also, in a classical singing performance, the tabla is always tuned such that the pitch perceived due to the Na stroke is the same as the tonic of the singer, resulting in the co-occurrence of voice and tabla partials in the same spectral region. Interference signals, simulated for use in the experiments in the following sections, are based on the aforementioned acoustic characteristics of the Na stroke.

3 Performance Evaluation of PDAs

From the last section, we see that the extended duration and complex tonal nature of the tabla strokes (such as ‘Na’ and ‘Tun’) cause the most significant degradation. Such strokes have spectra characterized by a few prominent harmonics. Since the melodic voice is also characterized by a harmonic pattern, it is hypothesized that a harmonic pattern matching algorithm may help to detect the melodic pitch. The two-way mismatch (TWM) PDA is such a spectral-domain method [7]. It has been widely applied to monophonic pitch estimation [8]. To investigate the specific advantage, if any, of a spectral domain PDA, we also evaluate a well-known correlation domain PDA based on the ACF on the same signals. Some modifications to the implementation of the basic ACF are proposed in [6], which increase its robustness to additive noise, large pitch ranges and rapidly changing sounds and decrease its sensitivity to strong formants. These modifications have been incorporated in our implementation.

In this section, we briefly review the TWM algorithm, and present a simulation experiment to compare pitch detection accuracy with that of the ACF PDA.

3.1 Two-way Mismatch Method

The two-way mismatch PDA [7] detects the pitch as that trial fundamental frequency that best explains the measured partials of the signal. A mismatch error is computed between a predicted harmonic spectral pattern and the spectral peaks detected in the signal. The trial fundamental frequency, in the selected search range, for which the best match between predicted and measured spectra is achieved, is indicated by the location of the global minimum of the TWM error.

The TWM error is computed as a weighted combination of two errors, one based on the frequency difference between each partial in the measured sequence and its nearest neighbour in the predicted sequence, and the other, based on the mismatch between each harmonic in the predicted sequence and its nearest neighbour in the measured partials. This two-way mismatch helps avoid octave errors in the absence of interference. The specific form of the error function (Eqs. 1-3 in [7]) applies an amplitude-weighted penalty to a normalized frequency error between measured and predicted partials for that trial fundamental frequency.

The presence of strong white noise interference can distort the magnitude spectrum leading to an incorrect selection of peaks, resulting in a distortion of the data input to the pitch computation procedure itself. However, whether the presence of a harmonic interference causes pitch detection errors largely depends on the actual number of interference harmonics detected in the magnitude spectrum as compared to the target (voice) harmonics.

Recommended TWM parameter values, empirically determined for quasi-harmonic signals [7], are $p = 0.5$, $q = 1.4$, $r = 0.5$ and $\rho = 0.33$. The most important parameter is ‘p’, higher values of which serve to emphasize low frequency region errors. Since we do not explicitly take into account prior knowledge of the frequency location of interference partials, we use a lower value of $p = 0.1$ in our application. Additionally, it is found that using $\rho = 0.25$ leads to better pitch detection performance due to the consequent higher emphasis on predicted-to-measured partials differences. This favours the target voice fundamental when the interference is characterized by a few partials only.

3.2 Experiment

Generation of Test Signals. In order to comprehensively evaluate the accuracies of the TWM and ACF PDAs, simulated signals representing the worst case characteristics of the vocal and percussion combination were generated. In order to simulate the large and rapid pitch modulations in Indian classical singing and also the typical vocal range of a singer, two vowels (/a/) are synthesized, at a sampling frequency of 22050 Hz, with time-varying fundamental frequency. The time variation smoothly sweeps ± 1 octave from a chosen base pitch. To simulate male and female voices, low (150 Hz) and high (330 Hz) values of base pitch are chosen respectively. The maximum change of the pitch, which takes place around the base pitch region, between two frame centers (spaced 10 milliseconds apart) is 5% of the base pitch. The simulated vowel has a length of 12 sec (an approximate upper limit on the duration of a typical singing spurt) in which the pitch undergoes 10 oscillations around the base pitch.

As for simulating the percussive interference, based on the spectral characteristics immediately after its onset, as discussed in Sec. 2, a *steady* complex tone having four partials in the ratio 2:3:4:5 for a fundamental frequency which is the same as the base pitch of the vowel, with the second partial being 5 dB above the remaining partials, is generated. The simulated vocal and interference signals are mixed at an SNR of 0 dB.

Simulation Results. For each of the PDAs under consideration, the detected pitch (computed once every 10 ms) was compared with the ground truth pitch of the simulated vocal signal. A gross error is defined to occur when the detected pitch is outside the ±6% neighbourhood of the ground truth pitch. Table 1 shows the percentage gross error across the signal duration (1257 pitch estimates). It is clear that in both cases the TWM PDA demonstrates greater robustness to the complex tonal interference than the ACF PDA.

In the absence of any interference the ACF PDA may make octave errors i.e. may incorrectly pick a multiple or sub-multiple of the true pitch as the final pitch estimate. However, the presence of a strong tonal interference, may suppress the periodicity of the target (in this case the voice) altogether in the correlation domain, in which case the ACF might incorrectly pick the pitch of the interference or its octave multiple as the final pitch estimate [9]. Further discussion regarding the advantages of TWM over ACF for this data is given in Section 4.1.

Table 1. Gross error rates for TWM and ACF for the synthetic target signals at low and high base pitches + Na simulation added at 0 dB SNR

Target synthetic signal	Gross error rate (TWM)	Gross error rate (ACF)
Vowel at low base pitch	20.7 %	80.7 %
Vowel at high base pitch	21.2%	73.3 %

4 Dynamic Programming Based Post-processing

Dynamic programming-based smoothing [4] is applied to continuously voiced segments. The operation of DP can be explained by considering a state space where, for a given frame (j), each state (p,j) represents a possible pitch candidate. Any pitch contour can be seen as the path ((p(1),1), (p(2),2),…, (p(j),j),…, (p(N),N)) through this state space where p(j) is the pitch estimate at the j^{th} frame and N is the total number of frames in the given voiced segment. The measurement cost is the cost incurred while passing through each state i.e. E(p,j) is the measurement cost incurred at frame j for candidate p. For the time evolution of pitch, a smoothness cost W(p,p') is defined as the cost of making a transition from state (p,j) to state (p',j+1) where p and p' can be any candidate values in successive frames only. A local transition cost T, is defined as the combination of these two costs over successive frames (Eq. 1).

$$T(p(j+1), p(j), j+1) = E(p(j+1), j+1) + W(p(j), p(j+1)) \ . \quad \textbf{(1)}$$

Finally, an optimality criterion to represent the trade off between the measurement and the smoothness costs is defined in terms of a global transition cost (S), which is the cost of a path passing through the state space, by combining local transition costs across a segment, as shown in Eq. 2, where N is the number of frames in the segment.

$$S = \sum_{j=1}^{N-1} T(p(j+1), p(j), j+1) \ . \quad \textbf{(2)}$$

The pitch contour with the least global transition cost among all possible pitch contours for a given singing spurt is chosen as the final pitch contour.

4.1 Measurement Cost Function

The measurement cost function represents the reliability of each pitch candidate. For ACF the candidates available for a given frame are all those frequencies corresponding to lag values at which a peak is detected in the ACF, ranked in descending order of the strength of the ACF peak. For TWM, the candidates are all those trial frequencies at which a local minimum is found in the TWM error curve, ranked in ascending order of the TWM error value.

For both ACF and TWM, the suitability for post-processing is determined by the quality of the measurement cost or the salience (reliability) of the underlying melodic (true) pitch. Salience of the candidate at the true pitch for both PDAs is computed as shown in Eq. 3,

$$\text{Salience} = 1 - \frac{\left(LS_{tr} - LS_{tp}\right)}{LS_{tr}} , \quad \textbf{(3)}$$

where LS_{tr} and LS_{tp} are the local strengths of the top-ranked and the true-pitch candidates respectively, as output by the PDA. The local strength of a candidate for ACF is computed as per [6], while the local strength of a candidate for TWM is 1 minus the TWM error value at that candidate frequency.

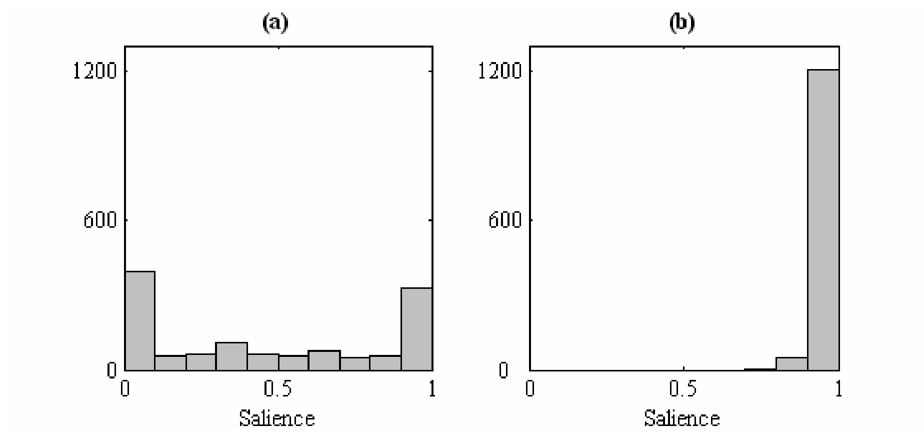

Fig. 3. Histogram of the salience of the candidate at true pitch for the target at low base pitch for **a.** ACF PDA and **b.** TWM PDA. Absence of a candidate at the true pitch is indicated by a salience value of 0.

Fig. 3 shows the histogram of salience values of the candidate at the true melodic pitch for each of the ACF and TWM PDAs, for all frames, for the Na stroke simulation added to the target at low base pitch at 0 dB SNR. A salience of 1 indicates that the true and top ranked pitch candidate are the same while a salience of 0 indicates that there is no candidate at the true pitch present in the list of candidates. It can be seen that for a large percentage of frames (27.5%) for ACF, the true pitch does not appear in the list of candidates at all. So the lowest possible gross error rate achievable after DP-based smoothing is 27.5%. On the other hand, a candidate at the true pitch is always present in the TWM output. Also, the salience of the true pitch candidate, for TWM, is always very high (never below 0.7), but for ACF, a large number of true pitch candidates have low values of salience. The results are similar for the target at high base pitch.

The high salience of the true pitch candidates for TWM, in the presence of a strong, pitched interference, is an indication of the robustness of TWM to such interference. When confronted with two simultaneously occurring pitched signals, the fundamental frequency of the signal with a larger number of detected harmonics will generally have a lower value of TWM error. In a separate simulation exercise, we found that gradually increasing the number of harmonics in the interference signal, keeping the overall target signal to interference power fixed, increases greatly the pitch detection errors. It also reduces, but to a lesser extent, the salience of the true pitch. Considering that the voice has a larger number of significant partials as compared to the tabla interference, the TWM PDA is able to detect the voiced pitch accurately or at most make octave errors. The ACF, on the other hand, is found to have a performance dependent on the relative power of the interference rather than on its spectral structure in terms of number of harmonics present. This is consistent with the known behaviour of the ACF PDA where the stronger periodic signal dominates the autocorrelation function when simultaneous signals of different periodicities are present.

The above results indicate that the TWM PDA shows significantly better potential in a pitch tracker based on DP post-processing for Indian classical music. Accordingly, the post-processor used, for all subsequent experiments, operates on the candidates output by the TWM PDA.

4.2 Smoothness Cost Function

The smoothness cost function should represent the knowledge that the vocal pitch contour is very likely to be smooth in some sense. Since the percussive interference is not continuous in time, smoothness constraints can play an important role in reconstructing the parts of the contour obscured by errors given that portions of the pitch contour are uncorrupted. In music, it is appropriate to define smoothness in terms of differences in log pitch. The smoothness cost function in the DP formulation of [6] is shown in Eq. 4

$$W(p,p') = \text{OctaveJumpCost}.\left|\log_2\left(p'/p\right)\right| \;, \tag{4}$$

where p and p' are the pitch estimates for the previous and current frames. The OctaveJumpCost (OJC) values tested are 0.35 and 1.0, corresponding to increasing penalties for the same pitch transitions.

With a view to determine a smoothness cost that is more musicological knowledge based, a distribution of inter-frame pitch transitions was obtained from pitch contours extracted from 20 minutes of continuous monophonic singing segments of two male and two female classical singers. The distribution, shown in Fig. 4, indicates that most pitch transitions are in a close neighbourhood, and the probability of a given transition decreases rapidly (but nonlinearly) with increasing magnitude. At larger magnitudes of pitch transition, the probability falls off very slowly to near zero. It is reasonable to base the smoothness cost on the probability of the pitch transition. A cost function that satisfies the above criteria is defined in Eq. 5.

$$W(p,p') = 1 - e^{\frac{-\left(\log_2(p') - \log_2(p)\right)^2}{2\sigma}} \quad . \tag{5}$$

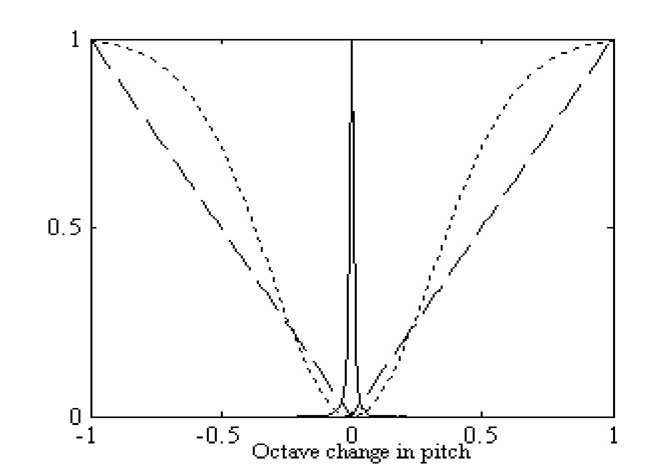

Fig. 4. Normalized distribution (*solid curve*) of log pitch transitions between adjacent frames (at 10 ms intervals) computed from true pitch contours of several songs sung by male and female singers. Log cost function (OJC =1) (*dashed curve*) and Gaussian cost function (σ = 0.1) (*dotted curve*) respectively.

For convenience, the cost functions defined in Eq. 4 and Eq. 5 are henceforth referred to as the log and Gaussian cost functions respectively. They are shown as dashed and dotted lines respectively in Fig. 4.

4.3 Experimental Evaluation of Different Smoothness Cost Functions

Synthetic Target and Synthetic Interference. DP-smoothing is carried out on the output of TWM for the two cases of synthetic vowels (at low and high base pitches) added to a steady complex tone (Na simulation) at 0 dB SNR (see Section 3.2). The different cost functions experimented with are the Log cost function with OJC = 0.35 and 1, and the Gaussian cost function with a standard deviation (σ) of 0.1.

From Table 2, we can see that the least number of errors is achieved for the Gaussian log cost function for both cases. An analysis of the errors revealed that errors are made for those frames that are around regions where the target pitch contour is close to the interference pitch. For these frames, the minima values at the interference pitch in the TWM error are close to the minima at the target pitch, if at all these are distinct. Thus the measurement cost for the interference and target pitches are very close. Pitch transitions in these regions lie within the interval where the distribution of pitch transitions is high (Fig. 4). As such, these transitions are penalized more by the log cost function than the Gaussian cost function, since the former is steeper in this region. As a result, the pitch contour estimated after DP with the log cost tends to 'lock on' to the interference pitch for short durations whenever the target pitch crosses the interference pitch, since a transition moving away from the interference pitch is penalized more. This is not the case for the pitch contour estimated by DP with the Gaussian cost function, which is able to track the target pitch contour quite accurately. Thus, the flatter nature of the Gaussian cost function, for pitch transitions ranging from -0.2 to + 0.2 octaves, as compared to the log cost function, results in fewer errors for DP smoothing.

Table 2. Gross error rates (GER) for TWM+DP for different smoothness cost functions for targets at low and high base pitch.

Cost Function	Parameter value	GER (low base pitch)	GER (high base pitch)
Log	OJC = 0.35	5.7 %	5.7 %
Log	OJC = 1.0	3.2 %	2.5 %
Gaussian Log	$\sigma = 0.1$	1.0 %	1.0 %

Synthetic Target and Real Interference. Real signals consisting of a rapid sequence of Tabla strokes, both impulsive and tonal, were added over the duration of the synthetic vowels at low and high base pitch, such that the average local SNR for each stroke onset was about -5 dB. The pitch tracker (TWM+DP) was tested using different cost functions on the resulting signals. Results are shown in Table 3. For the target at low base pitch, the log cost function (OJC = 1.0) yields the same result as the Gaussian cost function, but for the target at higher base pitch, the performance of the latter is marginally better. These results reinforce the case for using a probability-based transition cost function for DP.

Table 3. Gross error rates for TWM+DP for different smoothness cost functions for targets at low and high base pitch, and real interferences.

Cost Function	Parameter value	GER (low base pitch)	GER (high base pitch)
Log	OJC = 0.35	6.7 %	5.7 %
Log	OJC = 1.0	1.9 %	1.5 %
Gaussian Log	$\sigma = 0.1$	1.9 %	1.2 %

4.4 Test Case with a Real Signal

Fig. 5 compares the performance of ACF and TWM, with and without post-processing using the same smoothness cost function, for the real signal mentioned in Fig. 1. The pitch contour estimated by the ACF PDA (Fig. 5a) is degraded by all strokes but DP is able to recover from the impulsive strokes. However, the estimated pitch contour 'locks on' to the interference fundamental during a significant time period after the onset of the tonal stroke due to the low salience of the voice pitch in the presence of strong percussive interference. On the other hand, the TWM PDA estimated pitch contour (Fig. 5b) is not affected by the tonal tabla stroke at all in this case, and the solitary error is made due to the impulsive tabla stroke. This error is corrected by DP-based smoothing.

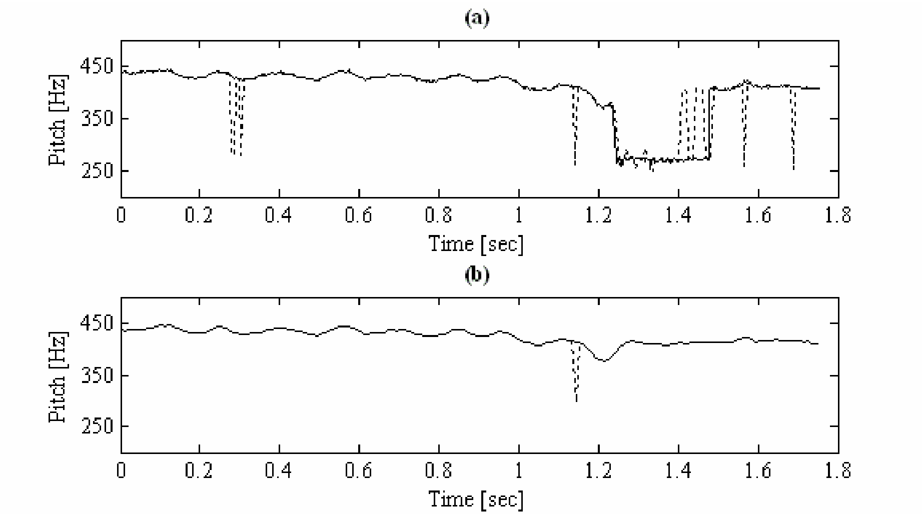

Fig. 5. Pitch contour for the real signal in Fig. 1 estimated by **a.** ACF and **b.** TWM, with (*solid curve*) and without (*dashed curve*) post-processing respectively.

5 Conclusion

The problem of pitch tracking in the presence of typical Indian percussive interference was considered. The salient characteristics of the tabla interference are that it is intermittent (not continuously present) and that its spectral structure comprises of strong, but few, harmonic components. A pitch tracker that uses the TWM PDA followed by DP-based smoothing was investigated on simulated and real signals. It is seen that the raw pitch estimate of the TWM PDA is accurate in the presence of the spectrally overlapping harmonic interference. This was attributed to the specific form of the TWM error function which emphasizes the spectral structure of the interference in terms of number of harmonics present and, to a lesser extent, the strengths of the harmonics. Even in unfavourable cases, when pitch detection errors occur, the target melodic notes remain salient in the PDA output. This allows DP-based post-processing to recover the pitch contour based on smoothness constraints. A new smoothing cost based on the observed distribution of pitch transitions was shown to be particularly effective in recovering the melodic pitch from gross errors.

From these results, two rules regarding the choice of measurement and smoothness cost functions for dynamic programming based smoothing of a pitch contour output by a PDA can be formulated. (1) The measurement cost must reflect high salience of a candidate at the melodic pitch, even in the presence of interference. (2) An effective smoothness cost function must be related to inter-frame pitch transition probabilities. Future work is targeted toward the wider testing of the proposed pitch tracking algorithm, and comparison with more recently proposed PDAs such as a bandwise ACF based PDA [10].

References

1. Krishnaswamy, A.: Pitch Measurements Versus Perception of South Indian Classical Music. Proc. of the Stockholm Music Acoustics Conference, Sweden (2003)
2. Subramanian, M.: An Analysis of Gamakams Using the Computer. Sangeet Natak, Vol. 37. (November 2002) 26-47
3. Battey B.: Bezier Spline Modeling of Pitch-continuous Melodic Expression and Ornamentation. Computer Music Journal, Vol. 28, No. 4. (2004) 25-39
4. Ney, H.: Dynamic Programming Algorithm for Optimal Estimation of Speech Parameter Contours. IEEE Trans. On Systems, Man and Cybernetics, Vol. SMC-13, No. 3. (April 1983) 208-214
5. Chordia, P.: Segmentation and Recognition of Tabla Strokes. International Conference on Music Information Retrieval (ISMIR). (2005) 107-114 (Database of Tabla Strokes at http://ccrma.stanford.edu/~pchordia/tablaStrokes/)
6. Boersma, P.: Accurate Short-term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound. Proc. of the Institute of Phonetic Sciences, Vol.17. (1993) 97-110
7. Maher, R., Beauchamp, J.: Fundamental Frequency Estimation of Musical Signals Using a Two-Way Mismatch Procedure. J. Acoust. Soc. Am., Vol. 95, No. 4. (April 1994) 2254-2263
8. Cano, P.: Fundamental Frequency Estimation in the SMS Analysis. Proc. of the Digital Audio Effects Workshop (DAFX). (1998)
9. Rao, P., Shandilya, S.: On the Detection of Melodic Pitch in a Percussive Background. J. Audio Eng. Soc., Vol. 52. (2004) 378-390
10. Klapuri, A., Virtanen, T., Holm, J.M.: Robust Multipitch Estimation for the Analysis and Manipulation of Polyphonic Musical Signals. Proc. of the Digital Audio Effects Workshop (DAFX). (2000)