PEDAGOGICAL TRANSCRIPTION FOR MULTIMODAL SITAR PERFORMANCE

Ajay Kapur, Graham Percival, Mathieu Lagrange, George Tzanetakis
Department of Computer Science
University of Victoria
British Columbia, Canada

ABSTRACT

Most automatic music transcription research is concerned with producing sheet music from the audio signal alone. However, the audio data does not include certain performance data which is vital for the preservation of instrument performance techniques and the creation of annotated guidelines for students. We propose the use of modified traditional instruments enhanced with sensors which can obtain such data; as a case study we examine the sitar.

1 INTRODUCTION

Historically, musical traditions were preserved via oral transmission. With the invention of written music, audio recordings, and video, more information can be retained. However, valuable performance data must still be passed by oral means. There will never be a technological replacement for face-to-face teaching, but new methods for archiving performance data will let us retain and disseminate more information.

Automatic music transcription is a well-researched area [3, 4, 5]. The novelty of our work is that we look beyond the audio data by using sensors to avoid octave errors and problems caused from polyphonic transcription. In addition, our work does not share the bias of most research that focuses only on Western music.

This paper describes a music transcription system for sitar performance. The sitar is a fretted stringed instrument from North India. Unlike many Western fretted stringed instruments (classical guitar, viola de gamba, etc) sitar performers pull (or "bend") their strings to produce higher pitches. In normal performance, the bending of a string will produce notes as much as a fifth higher than the same fret-position played without bending. In addition to simply showing which notes were audible, our framework also provides information about how to produce such notes. A musician working from an audio recording (or transcription of an audio recording) alone will need to determine which fret they should begin pulling from. This can be challenging for a skilled performer, let alone a beginner. By representing the fret information on the sheet music, sitar musicians may overcome these problems.

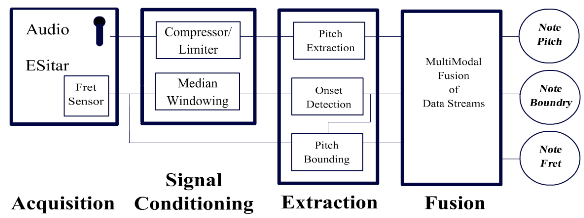

Figure 1. Block diagram of the system

2 METHOD

The Electronic Sitar (ESitar) [2] is an instrument which gathers gesture data from a performing artist. For the research described in this paper, fret data was captured by a network of resistors connecting each fret.

The fret sensor is translated into MIDI pitch values based on equivalent resistance induced by left hand placement on the neck of the instrument. Each fret has a "bucket" of values, converting raw sensor data into discrete pitch values seen in figure 2. Data was recorded at a sampling rate of $(44100 \div 512)$ Hz. The synchronized audio signal was recorded with a Shure Beta-57 microphone at 44100 Hz. The entire system is displayed in Figure 1.

2.1 Audio Signal Chain and Pitch Extraction

A compressor/limiter was used on the audio signal to generate a balanced amplitude for all frequencies. Without this step, our experiments yielded poor results for notes played at lower frequencies.

To automatically determine the pitch an implementation of the method described in [1] was used. We utilize the autocorrelation function to efficiently estimate the fundamental frequency ($f0$). For a time signal $s(n)$ that is stationary, the autocorrelation $r_s(\tau)$ as a function of the lag τ is defined as

$$r_s(\tau) = 1/N \sum_{j=t}^{t+N} s(t)s(t+\tau) \quad (1)$$

This function has a global maximum for $\tau = 0$. If there are also additional global maxima, the signal is called periodic and there exists a lag τ_0, the period, so that all these

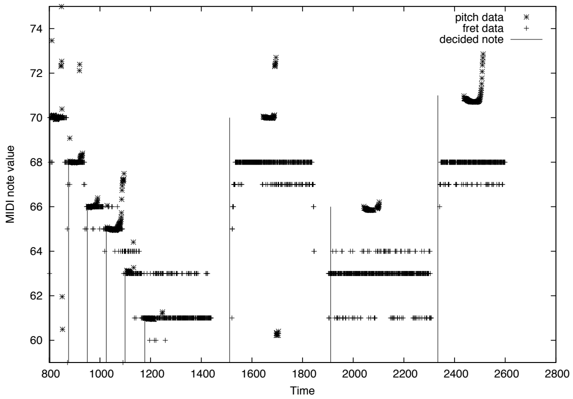

Figure 2. Fret data, audio pitches, and the resulting detected notes. The final three notes were pulled.

maxima are placed at the lags $n\tau_0$, for every integer n, with $r_s(n\tau_0) = r_s(0)$.

The inverse of the lag τ_0 provides an estimation of the fundamental frequency $f0$. The period is determined by scanning $r_t(\tau)$, starting at zero, and stopping at the first global maximum with non-zero abscissa. Quadratic interpolation is used to further improve the frequency estimation. In practical cases, the relative amplitude of those maxima may change and some others maxima may appear due to small aperiodicities of the signal. The issue is then to relevantly select which maximum corresponds to the $f0$ by considering several candidates under a plausible range and pick the one with the highest confidence, see [1] for further references on the algorithm.

2.2 Fusion with Fret Signal Chain

To compensate for noisy fret data, we smoothed the fret data samples comparing the median value of the previous ten samples with the median of the next ten samples (including the current sample). If the median values differed by more than a certain amount, we marked that sample as being a note boundary.

To get an accurate final result, pitch information from the audio signal chain is fused with onset and pitch boundaries calculated from the fret signal chain. The fret provided convenient lower and upper bounds on the pitch: a note cannot be lower than the fret, nor higher than a fifth (ie 7 MIDI notes) above the fret. Using the note boundaries derived from the fret data, we find the median value of the pitches inside the boundaries supplied by the fret data. These are represented by the vertical lines in Figure 2, and are the note pitches in the final output.

3 SHEET MUSIC

Although there are many computer notation programs for Western music, there are no such counterparts for Indian music. Indian notation is not standardized and there is no way to notate both frets and audible notes, so we invented

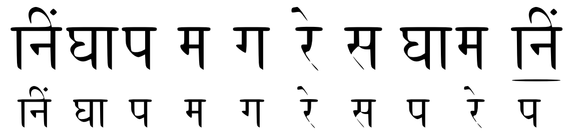

Figure 3. Sheet music of sitar performance, generated from the data used in Figure 2. The top *swara* are the audible notes, while the lower *swara* are the fret positions. The final three notes were pulled, so the audio and fret *swara* are different.

our own notation. In North Indian classical music, notes are described by seven *swaras*. They are known as *Shadja* (Sa), *Rishab* (Re), *Gandhar* (Ga), *Madhyam* (Ma), *Pancham* (Pa), *Dhaivat* (Dha), and *Nishad* (Ni). These are equivalent to the Western solfege scale. We produce sheet music (Figure 3) showing sitar musicians the audible note played and which fret was used.

4 CONCLUSION

We have developed a system which produces sheet music for sitar musicians. Although the system works well for simple sitar melodies, it currently does not detect glissandi and bends. In addition, since we detect note boundaries by examining the fret data, we cannot detect multiple notes plucked on the same string. Future work in this area is planned, in addition to experimenting with other pitch detection algorithms (such as the YIN method). Once we can reliably detect typical sitar flourishes (glissandi and note bends), we must invent notation to display such expressive marks.

5 REFERENCES

[1] P. Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proc. of the Institute of Phonetic Sciences, Amsterdam*, volume 17, pages 97–110, 1993.

[2] A. Kapur, A. J. Lazier, P. Davidson, R. S. Wilson, and P. R. Cook. The electronic sitar controller. In *NIME*, pages 7–12, 2004.

[3] A. P. Klapuri. Automatic music transcription as we know it today. *Journal of New Music Research*, 33(3):269–282, 2004.

[4] A. Loscos, Y. Wang, and W. J. J. Boo. Low level descriptors for automatic violin transcription. In *Proc. of ISMIR 2006*, pages 164–167, October 2006.

[5] J. Yin, T. Sim, Y. Wang, and A. Shenoy. Music transcription using an instrument model. *Proc. ICASSP '05*, 3, March 2005.