Recognition and Prediction in a Network Music Performance System for Indian Percussion

Mihir Sarkar
The Media Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA
mihir@media.mit.edu

Barry Vercoe
The Media Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA
bv@media.mit.edu

ABSTRACT

Playing music over the Internet, whether for real-time jamming, network performance or distance education, is constrained by the speed of light which introduces, over long distances, time delays unsuitable for musical applications. Current musical collaboration systems generally transmit compressed audio streams over low-latency and high-bandwidth networks to optimize musician synchronization. This paper proposes an alternative approach based on pattern recognition and music prediction. Trained for a particular type of music, here the Indian tabla drum, the system called *TablaNet* identifies rhythmic patterns by recognizing individual strokes played by a musician and mapping them dynamically to known musical constructs. Symbols representing these musical structures are sent over the network to a corresponding computer system. The computer at the receiving end anticipates incoming events by analyzing previous phrases and synthesizes an estimated audio output. Although such a system may introduce variants due to prediction approximations, resulting in a slightly different musical experience at both ends, we find that it demonstrates a high level of playability with an immediacy not present in other systems, and functions well as an educational tool.

Keywords

network music performance, real-time online musical collaboration, Indian percussions, tabla bols, strokes recognition, music prediction

1. INTRODUCTION

The main challenge in playing music in real-time over a computer network is to overcome the time delay introduced by the network's latency which is bounded by the speed of light. For an online musical performance system to work, musicians have to be perceptually synchronized with one another while data travels between two locations. We solve this problem by developing a program that (i) recognizes individual drum strokes and identifies standard drumming primitives from the input signal, (ii) transmits symbolic events over the network instead of an audio stream, and (iii) synthesizes rhythmic phrases with the appropriate pitch and tempo at the output by using previous events to predict current patterns.

We present here *TablaNet*, a real-time musical collaboration system for the tabla that involves machine listening. We selected the tabla, a percussive instrument widely used in North India, not only because of our familiarity with it, but also because of its "intermediate complexity": although tabla compositions are based only on rhythmic patterns without melodic or harmonic structure, different strokes can produce a variety of more than 10 pitched and unpitched sounds called *bols* which contribute to the tabla's expressive potential. We expect our results to generalize to other percussion instruments of a similar nature. This work has resulted in a playable prototype and a simulation environment for testing and demonstration.

In this paper we briefly survey previous work relevant to this project. Then we outline our approach by describing the system architecture and some of the design choices. In particular we focus briefly on the tabla strokes recognition module.

This paper is accompanied by a live demonstration of the system where a tabla player interacts with the system which identifies tabla strokes in real-time, transmits rhythmic phrases over a simulated network channel and recreates patterns at the other end.

2. RELATED WORK

2.1 Network Music Performance

Online music collaboration has been and continues to be the source of several commercial endeavors (from the defunct Rocket Network to Ninjam and Lightspeed Audio Labs, a new startup still in stealth mode). However, efforts towards distributed music performances (see for example [16]) are constrained by the inherent latency of computer networks. While some projects have attempted to stream uncompressed audio on the Internet (e.g. [9]), most endeavors minimize the time delay by sending MIDI events or using audio compression with reasonable algorithmic complexity (e.g. [19], [5], [12]). As an alternative, some projects target local area networks and ad-hoc wireless mesh networks [22] to obviate the issue of latency, and others use improved and faster networks, such as the experimental Internet2 (e.g. [1] and Sawchuk2003). In addition, researchers have found

NIME07, June 7-9, 2007, New York, NY

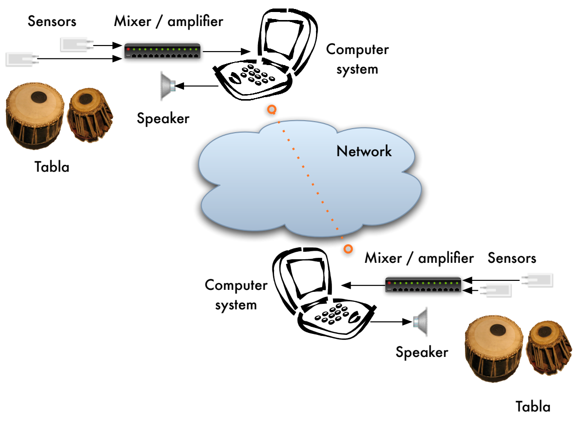

Figure 1: TablaNet System Diagram

other creative ways to interact musically over the Internet, for example by converting transmission delays into reverberation [3], or by developing novel musical interfaces [24]. Another approach is to increase the time delay (instead of trying to fight it) to a musically relevant quantity, such as the ”one-phrase delay” introduced by Goto [11] and implemented by Ninjam. Studies have also been conducted on the effects of time delay on musician synchronization (see [4], [7] and [20]), paving the way for roadmaps in the area of networked musical performance, such as [17] and [23].

2.2 Tabla Analysis & Synthesis

For a description of the tabla, the reader is invited to see, for instance, [15]. As one of the most popular Indian instruments, the tabla has a complex timbral quality which includes both pitched and unpitched sounds. Several researchers have attempted to analyze tabla sounds as well as synthesize them. In fact there have been a number of studies to recognize tabla strokes using statistical pattern classification (from [10] and [6] to [21] and [8]). However these methods analyze performances recorded in controlled environments, whereas our system is built for live performance settings where drum strokes are captured using sensors other than microphones to avoid feedback and ambient noise. Moreover, different types of electronic tabla controllers (see [13] and [15]) have been developed, some of which use tabla sounds generated by physical models [14]. On the representation side, the *Bol Processor* [18] implements a linguistic model to describe complex rhythms.

3. PROPOSED APPROACH

3.1 System Design

The TablaNet system architecture is described in Figure 1. At the near-end, a pair of sensors (one for each drum) captures the strokes that are played on the tabla. The signals from both the sensors are mixed and pre-amplified, and sent to the Analog-to-Digital converter on the near-end computer. After processing the input audio signal, the computer sends symbols over the network to a far-end computer installed with the same software. The receiving computer interprets the events transmitted in the symbols and generates an appropriate audio output. The system is symmetrical and full duplex so that each musician can simultaneously play, and listen to the musician at the other end.

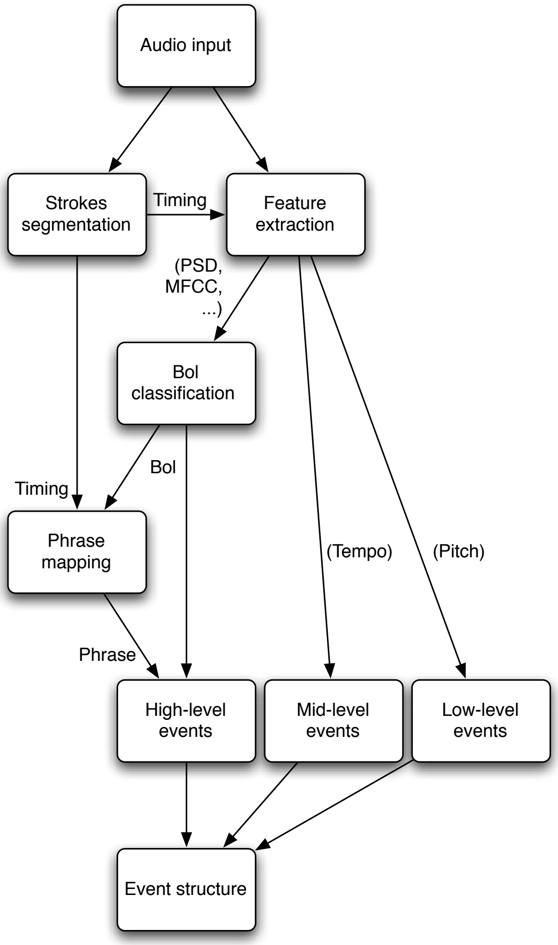

Figure 2: Transmitter Software Block Diagram

3.2 Hardware Interface

To avoid feedback from the speakers into a microphone, we use piezoelectric vibration sensors pasted directly on the tabla heads with thin double-sided tape. The output of these sensors is fed into a pre-amplified mixer so that the resulting monophonic signal can be connected to the microphone input on the target computer. The reason for this is that many low-cost computers (i.e. the $100 laptop [22]) may only have a monophonic input. The fact that the sounds coming from both drums are mixed is not an issue because strokes that are played on the right drum (*dayan*) are distinct from those played on the left drum (*bayan*), and from those played on both drums simultaneously (see Table 1).

Each computer has an external or built-in amplified speaker to play the audio output estimated from the other end.

3.3 Software Architecture

We call the analysis modules, which convert the incoming audio signal to symbols, the Transmitter. On the other side, the Receiver contains the modules that listen to the network and convert incoming symbols back to an audio

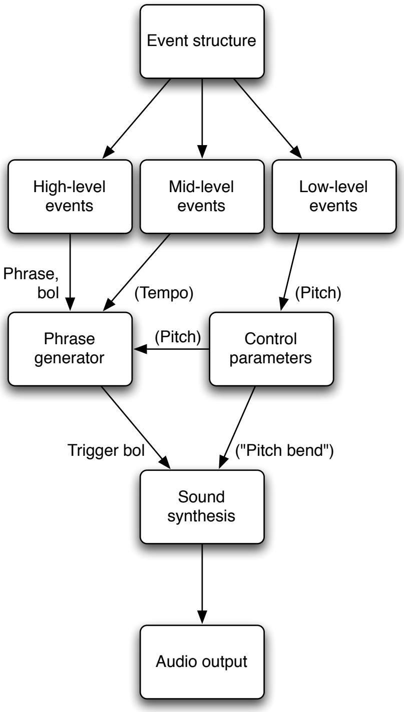

Figure 3: Receiver Software Block Diagram

output. Both Transmitter and Receiver are present on the near-end and the far-end computers.

The software block diagram of the Transmitter is presented in Figure 2. As a first step, individual drum sounds are isolated by detecting onsets (amplitude and phase discontinuities) and stored into a frame. Then, audio features are extracted from each frame. These features include spectral domain components and cepstral domain coefficients for bol recognition, pitch tracking for general tabla tuning and occasional pitch slides (on the bayan), and tempo data computed from the onset timings. The bol classifier runs on each frame containing a stroke, and outputs its result to a phrase mapping algorithm that selects the most probable rhythmic phrase among those learned previously by matching them with the incoming stream of bols. These events are combined hierarchically into a data structure, and sent asynchronously over the network as they become available.

When such a data structure reaches the Receiver (Figure 3) through the network, the various events are demultiplexed. High-level events like phrases and individual bols influence the phrase generator, which estimates the rhythmic pattern to be output locally. This module keeps track of previous bols and phrases, as well as tempo changes and pitch bends, in order to adapt dynamically to the far-end musician's playing style. Based on the sequence of bols, the phrase generator triggers samples in the sound synthesis engine at the appropriate time. Low-level events such as pitch variations control parameters of the sound synthesis engine.

3.4 Tabla Strokes Recognition

The system recognizes the tabla strokes listed in Table 1 by analyzing the audio signal using pattern recognition algorithms. Initial features included Power Spectral Density (Periodogram method), Mel-Frequency Cepstrum Coefficients. Preliminary results using a 512-bin FFT-based PSD with a fixed window length of 250ms reduced to 8 coefficients with Principal Component Analysis gives above 70% recognition rate with the k-Nearest Neighbors (k=3) algorithm. The confusion matrix provides results close to human perception (errors appear in strokes that are also difficult to distinguish for a human listener familiar with the instrument). Later results, with recognition rates above 95%, use the constant Q transform [2], a logarithmically-scaled DFT, and time slices within each frame.

3.5 Network Transmission

Tempo and a symbolic representation of rhythmic events are transmitted to the far-end computer over a non-guaranteed connectionless User Datagram Protocol network layer. The UDP protocol has been proven to perform better than TCP over IP for this application.

3.6 Music Prediction and Synthesis

Musicians at both ends play to a common metronome. Initially, each system plays a predefined phrase determined by the interaction mode (i.e. call-and-reponse, accompaniment). Upon receiving data asynchronously, the system dynamically alters its output parameters (e.g. stroke sequence, ”groove”, tempo) by matching the new data structure to a set of known (hard-coded) and learned (based on the interaction history) musical patterns in a particular metric (or *taal*). This is done using Dynamic Bayesian Networks. Tabla sounds are then synthesized using sample playback with parameters that allow effects such as pitch bends.

4. EVALUATION

The system presented here achieves a player-independent tabla strokes recognition rate above 95%, which is comparable with perceptual recognition results for an experienced musician.

The round-trip latency over the Internet can range from less than a millisecond on a LAN to almost half-a-second between some countries—not counting the system's algorithmic delay. These values, above the generally accepted 20ms threshold for musician synchronization, support the need for this system, especially with asynchronous networks.

For qualitative results, preliminary user studies have been performed with tabla players of various levels (beginner, less than 1 year experience; intermediate, from 1 to 3 years experience; and expert, more than 3 years experience). Experiments involved activities in the areas of:

- Distance learning (between a teacher and a student)
- Rhythmic accompaniment (two musicians playing simultaneously)
- Call and response (called *Jugalbandi*)

Network latency was simulated using median and worst case figures. After playing on the system for various periods of time, either with another musician or with the simulator, tabla players were asked to comment on whether the system met their expectations in terms of ”playability” (variety, quantization artifacts, sound quality). Responses were collected in a survey. Initial results showed that musicians had

Table 1: Tabla Strokes

Bol	Na	Tin	Ga	Ka	Dha	Dhin	Te	Re	Tat	Thun
Dayan (right drum)	✓	✓			✓	✓	✓	✓	✓	✓
Bayan (left drum)			✓	✓	✓	✓				

the impression of playing with another musician or with a "musically intelligent machine", rather than with a machine simply playing back recorded sequences.

Other proposed evaluation schemes include a "Musical Turing test" where players are asked to play either with another musician or with the computer system without knowing which one is at the other end, and then guessing, after a period of interaction, which one was actually at the other end.

5. CONCLUSIONS

This networked tabla performance system creates a real-world musical interaction between two tabla players over a computer network. We implemented a continuous tabla strokes recognizer, and developed a real-time tabla phrase prediction engine. This study suggests that although playing on a system such as TablaNet with a musician located across the network does not necessarily provide an experience identical to playing with a musician located in the same room, it creates new opportunities for learning and entertainment. Further work includes supporting more than two players playing together, as well as adding other types of musical instruments. We also propose to use a physical model for the sound synthesis engine in order to achieve higher control over the sound (especially in regard to pitch slides).

We hope that this work will be carried forward by others who wish to further and implement the principles presented here to other instruments and musical styles.

6. ACKNOWLEDGMENTS

We would like to thank Prof. Rosalind Picard for her encouragements, and the Music, Mind and Machine group members at the MIT Media Lab for their feedback and ideas.

7. REFERENCES

[1] R. Bargar, S. Church, A. Fukuda, J. Grunke, D. Keislar, B. Moses, B. Novak, B. Pennycook, Z. Settel, J. Strawn, et al. AES white paper: Networking audio and music using Internet2 and next-generation Internet capabilities. Technical report, AES: Audio Engineering Society, 1998.

[2] J.C. Brown. *Calculation of a Constant Q Spectral Transform*. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1990.

[3] C. Chafe. Distributed Internet Reverberation for Audio Collaboration. In *AES (Audio Engineering Society) 24th Int'l Conf. on Multichannel Audio*, 2003.

[4] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan. Effect of Time Delay on Ensemble Accuracy. In *Proceedings of the International Symposium on Musical Acoustics*, 2004.

[5] A. Chatwani and A. Koren. Optimization of Audio Streaming for Wireless Networks. Technical report, Princeton University, 2004.

[6] A.A. Chatwani. Real-Time Recognition of Tabla Bols. Princeton University, Senior Thesis, May 2003.

[7] E. Chew, R. Zimmermann, A.A. Sawchuk, C. Kyriakakis, C. Papadopoulos, ARJ François, G. Kim, A. Rizzo, and A. Volk. Musical Interaction at a Distance: Distributed Immersive Performance. In *Proceedings of the MusicNetwork Fourth Open Workshop on Integration of Music in Multimedia Applications, September*, pages 15–16, 2004.

[8] P. Chordia. Segmentation and Recognition of Tabla Strokes. In *Proc. of ISMIR (International Conference on Music Information Retrieval)*, 2005.

[9] J.R. Cooperstock and S.P. Spackman. The Recording Studio that Spanned a Continent. In *Proc. of IEEE International Conference on Web Delivering of Delivering of Music (WEDELMUSIC)*, 2001.

[10] O.K. Gillet and G. Richard. Automatic Labelling of Tabla Signals. In *Proc. of the 4th ISMIR Conf.*, 2003.

[11] M. Goto, R. Neyama, and Y. Muraoka. RMCP: Remote Music Control Protocol—Design and Applications—. *Proc. International Computer Music Conference*, pages 446–449, 1997.

[12] X. Gu, M. Dick, U. Noyer, and L. Wolf. NMP-a new networked music performance system. In *Global Telecommunications Conference Workshops, IEEE*, pages 176–185, 2004.

[13] J. Hun Roh and L. Wilcox. Exploring Tabla Drumming Using Rhythmic Input. In *CHI'95 proceedings*, 1995.

[14] A. Kapur, P. Davidson, P.R. Cook, P. Driessen, and A. Schloss. Digitizing North Indian Performance. In *Proceedings of the International Computer Music Conference*, 2004.

[15] A. Kapur, G. Essl, P. Davidson, and P.R. Cook. The Electronic Tabla Controller. *Journal of New Music Research*, 32(4):351–359, 2003.

[16] A. Kapur, G. Wang, P. Davidson, PR Cook, D. Trueman, TH Park, and M. Bhargava. The Gigapop Ritual: A Live Networked Performance Piece for Two Electronic Dholaks, Digital Spoon, DigitalDoo, 6 String Electric Violin, Rbow, Sitar, Table, and Bass Guitar. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME), Montreal*, 2003.

[17] A. Kapur, G. E. Wang, P. Davidson, and P. R. Cook. Interactive Network Performance: a dream worth dreaming? *Organised Sound*, 10(03):209–219, 2005.

[18] J. Kippen and B. Bel. Computers, Composition and the Challenge of "New Music" in Modern India. *Leonardo Music Journal*, 4:79–84, 1994.

[19] J. Lazzaro and J. Wawrzynek. A case for network musical performance. In *Proceedings of the 11th international workshop on Network and operating systems support for digital audio and video*, pages 157–166. ACM Press New York, NY, USA, 2001.

[20] T. Mäki-Patola. Musical Effects of Latency. *Suomen Musiikintutkijoiden*, 9:82–85, 2005.

[21] K. Samudravijaya, S. Shah, and P. Pandya. Computer Recognition of Tabla Bols. Technical report, Tata Institute of Fundamental Research, 2004.

[22] B. Vercoe. Erasing the Digital Divide: Putting your Best Idea on the $100 Laptop. Keynote lecture, WORLDCOMP'06, Las Vegas, June 2006.

[23] G. Weinberg. Interconnected Musical Networks: Toward a Theoretical Framework. *Computer Music Journal*, 29(2):23–39, 2005.

[24] G. Weinberg. Local Performance Networks: musical interdependency through gestures and controllers. *Organised Sound*, 10(03):255–265, 2005.