

Exploring Tabla Drumming Using Rhythmic Input

jae Hun Roh

MIT Department of EECS
1 Amherst St., E40-218 RPCP
Cambridge, MA 02139
Tel: 1-617-253-6828
E-mail: jhroh@mit.edu

Lynn Wilcox

Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
Tel: 1-415-851-2694
E-mail: wilcox@parc.xerox.com

ABSTRACT

We describe a system that enables the use of rhythmic input for exploring Indian *tabla* drumming. Rhythms drummed by the user on a pair of drum pads are mapped to tabla phrases using a hidden Markov model based recognizer. The recognized tabla phrases are played back to the user, while an animated visual representation of the phrase is displayed.

KEYWORDS: Multi-media, Tactile or gestural I/O, Auditory I/O, Intelligent Systems, Educational applications, Music applications.

INTRODUCTION

The term *multimedia* has come to be applied to systems which make use of multiple *presentation* media. However, the channel through which users access the contents of these presentations remains, in most contemporary systems, strictly *mono-media*. Here, we focus on the subset of multimedia systems which provide access to databases of stored data, rather than to systems which enable communications with other users. Interacting with these systems, often called *hypermedia* systems, typically involves selecting visual icons or keywords using a pointing device. In this work, users learn about drumming by doing drumming.

We propose an expansion of the notion of multimedia to include multiple input media. Specifically, we have developed a system for naive users to learn some basic aesthetic ideas of tabla drumming as it is practiced in North India. Our system responds to rhythmic "meta-tabla" input drummed out by the user on two pressure-sensitive drum pads. Meta-tabla is a partial specification of the tone and rhythm of a phrase in tabla drumming. We use a hidden Markov model (HMM) [5] of tabla phrases to build a recognizer which maps the user's meta-tabla to a complete phrase which might occur in tabla drumming. This work is part of an ongoing project to provide multimedia access to multimedia arts [4].

The use of multimedia input as a method for accessing speech, text, and image data was discussed in [2]. In [6], a

spoken keyword was used as input to a wordspotting system which located instances of the keyword in fluent speech. Computer music systems typically provide access to audio musical data by associating keywords or icons with discrete aesthetic units of a musical stream such as melodic motifs. In contrast to these systems, traditional and electronic instruments map the physical gestures of skilled users directly to musical expressions [3]. Our goal is to provide novice users with access to musical content through musical input.

TABLA DRUMMING

A tabla set consists of a pair of tuned drums with closed bottoms. The *tabla*, played with the right hand, and the larger *baya*, played with the left hand, might be considered respectively as treble and bass drums. A skilled tabla player can vary the pitch and timbre of the drum strokes by varying the pressure, damping, and shape of the hands on the drum. In the oral tradition of tabla, musicians communicate compositions to each other using onomatopaeic syllables, called *bols*, to represent different drum strokes. For example, *ta*, *te*, *ti*, and *tun* are played on the tabla, *ket* and *ge* are played on the baya. *Dha* and *dhin* are compound strokes played with both hands. Thus compositions can be sung as well as played.

The sequence of bols in a composition is constrained aesthetically by a variety of implicit rules and conventions. The sequence of bols in a phrase of tabla reflects a stylistic tradition; a given sequence may be common or rare (and stylistically incorrect). Further constraints on the bols reflect the *tala*, the metric framework of the composition.

VISUAL AND AUDIO INTERACTION

The user sees a pair of drums for input, a video display monitor, and audio speakers. The display presents both circular and linear timelines of the interaction. The meter is marked with an animated beat marker and an audible metronome click. (See Figure 1.) In time with the beat, the user taps out a rhythmic pattern on the two drums. The system echoes the rhythm as an unembellished rhythmic phrase, and marks the drum beats on the rhythmic cycle display. Then, using the tabla phrase recognizer, the system transforms the rhythm into a grammatically correct tabla pattern. This pattern is then displayed, played back to the user as drum sounds, and sung back as verbal bols. Throughout the duration of this interaction, the display offers a persistent visual record of the interaction, and graphically shows the correspondence between the user's input rhythm and the system's output.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of ACM. To copy otherwise, or to republish, requires a fee and/or specific permission.
CHI' Companion 95, Denver, Colorado, USA
© 1995 ACM 0-89791-755-3/95/0005...\$3.50

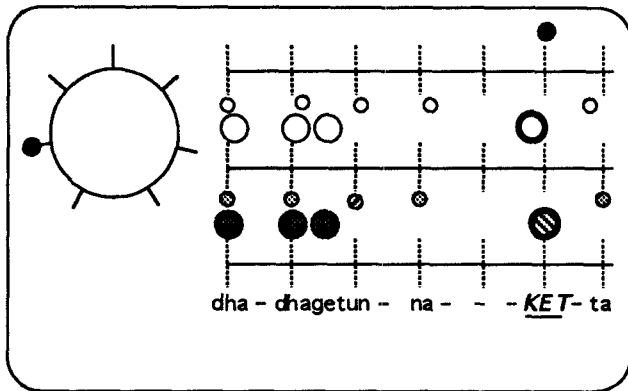


Figure 1. Visual Display and Sample Interaction

Figure 1 shows an sample interaction. The user taps out the sequence: B-BLR-R---L-R (i.e., both hands, pause, both, left, right,...: the rhythm is the same as "Shave and a haircut - two bits".) The system's response would be *dha - dha getun - na - - - ket - ta* .

The current beat is marked by the solid ball on the top row and on the circle. The user's input is marked on the second row, with left hand hits marked by the large circle, and right hand hits marked by the small circle. The system output is displayed on the next row. The shading indicates timbre, and the bols are displayed as text below. For example, *dha* is a compound stroke composed of *ge* and *na*. The current beat is highlighted in each line by a dark outline. Playback repeats cyclically until the user taps a new phrase. By varying the input rhythm, the user can explore variations of tabla phrasing.

TABLA PHRASE RECOGNITION

The user's hits are sensed by a commercial drum pad sensor and sent to a Macintosh computer as a serial MIDI (Musical Instrument Digital Interface) signal. The input to the tabla phrase recognizer includes the duration of each hit, that is, the time between the onset of the current hit and the next hit, and the identification (ID) of the drum pad.

The recognizer finds the tabla phrase which most closely matches the drummed input. We assume that tabla phrases can be modeled by a regular grammar [1], represented as a finite state automaton. The drummed input is modeled using a hidden Markov model [5]. Each state in the model is labeled with a tabla bol and an output duration. The allowable sequences of states (bols) correspond to "grammatical" tabla phrases.

The hidden Markov model defines the probability of an input sequence for any possible state sequence in the model. Each state has an output probability distribution over the drum ID and hit duration. Transition probabilities between states model the likelihood of various bol sequences. The output of the recognizer is the bol sequence which generates the observed drum input with the highest probability. The

recognizer finds this sequence using the Viterbi algorithm [5]. Note that this state sequence includes both bols and durations, so timing quantization of the output can be performed implicitly in the recognizer.

FIELD TRIALS

We conducted informal field trials of the system at the Exploratorium in San Francisco and at the Children's Museum in Boston. Although users found the work engaging, we found that the system presently lacks the depth to reward sustained exploration. Further collaborations with artists are planned to enrich the model with additional content, and to frame tabla within the larger context of Indian music and arts.

FUTURE WORK

We have developed a probabilistic model of tabla drumming which enables a novel system for *generating* phrases using rhythmic input. We are currently investigating the use of this model as a means to retrieve user specified tabla phrases from a large database of compositions.

ACKNOWLEDGMENTS

This work was made possible by the support of Ranjit Makkuni, Jan Pedersen, and others at Xerox PARC; Kathryn Vaughn and Barry Vercoc at the MIT Media Lab; Swapna Chaudhuri and Ravi Gutala at the Ali Akbar College of Music; Tom Humphrey and the staff at the Exploratorium; and Peggy Monahan at the Children's Museum.

REFERENCES

1. Bel, B. and Kippen, J. Bol Processor Grammars, in *Understanding Music with AI*. AAAI Press (1992) pp. 366-400.
2. Chen, F., Hearst, M., Kupiec, J., Pedersen, J., and Wilcox, L. Meta-data for Mixed Media Access, to appear in *SIGMOD Record Special Issue on Metadata for Digital Media*, (Dec. 1994).
3. Machover, T., Chung, J. *HyperInstruments: Musically Intelligent and Interactive Performance and Creativity Systems*. Intl. Computer Music Conf. (Columbus, Ohio, 1989)
4. Makkuni, R. *Museum of the Future, N-Dimensional: Project Gita-Govinda*. Xerox PARC Technical Report (1992).
5. Rabiner, L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* 77, 2 (Feb. 1989), pp. 257-285.
6. Wilcox, L., Smith, I., and Bush, M. Wordspotting for Voice Editing and Audio Indexing. *Proc. CHI'94* (May 1992), ACM Press pp. 655-656.