Integrating HyperInstruments, Musical Robots & Machine Musicianship for North Indian Classical Music

Ajay Kapur[1,2,3] Eric Singer[2] Manjinder S. Benning[1,3] George Tzanetakis[1] Trimpin

ajay@karmetik.com e@ericsinger.com manj@karmetik.com gtzan@cs.uvic.ca trimpin@att.net

[1]Music Intelligence and Sound Technology Interdisciplinary Centre (MISTIC), University of Victoria, BC, Canada

[2]League of Electronic Music Urban Robots (LEMUR), Brooklyn, NY, USA

[3]KarmetiK Technology (A Division of KarmetiK LLC), Reno, NV, USA

ABSTRACT

This paper describes a system enabling a human to perform music with a robot in real-time, in the context of North Indian classical music. We modify a traditional acoustic sitar into a hyperinstrument in order to capture performance gestures for musical analysis. A custom built four-armed robotic Indian drummer was built using a microchip, solenoids, aluminum and folk frame drums. Algorithms written towards "intelligent" machine musicianship are described. The final goal of this research is to have a robotic drummer accompany a professional human sitar player live in performance.

Keywords

Musical Robotics, Electronic Sitar, Hyperinstruments, Music Information Retrieval (MIR).

1. INTRODUCTION

There are many challenges in interfacing a human with a mechanical device controlled by a computer. Many methods have been proposed to address this problem, usually including sensor systems for human perception and simple robotic mechanisms for actuating the machine's physical response. Conducting this type of experiments in the realm of music is obviously challenging, but fascinating at the same time. This is facilitated by the fact that music is a language with traditional rules, which must be obeyed to constrain a machine's response. Therefore the evaluation of successful algorithms by scientists and engineers is feasible. More importantly, it is possible to extend the number crunching into a cultural exhibition, building a system that contains a novel form of artistic expression which can be used on stage.

The goal of this research is to make progress towards a system for combining human and robotic musical performance. We believe that in order for such a system to be successful it should combine ideas from the frequently separate research areas of music robotics, hyperinstruments, and machine musicianship.

The art of building musical robots has been explored and developed by musicians and scientists [1-6]. A recent review of the history of musical robots is described in Kapur [7]. The development of hyperinstruments plays a crucial role in obtaining data from the human's performance. Work and ideas by Machover [8] and Trueman [9] have greatly influenced the development of the interface and sensors described in this paper. The area of machine musicianship is another part of the puzzle. Robert Rowe (who also coined the term machine musicianship) describes a computer system which can analyze, perform and compose music based on traditional music theory [10]. Other systems which have influenced the community in this domain are Dannenberg's score following system [11], George Lewis's Voyager [12], and Pachet's Continuator [13].

There are few systems that have closed the loop to create a real live human/robotic performance system. Audiences who experienced Mari Kimura's recital with the LEMUR GuitarBot [14] can testify to its effectiveness. Gil Weinberg's robotic drummer Haile [15] continues to grow in capabilities to interact with a live human percussionist [16].

This paper describes a human-robot performance system based on North Indian classical music, drawing theory from ancient tradition to guide aesthetic and design decisions. Section 2 describes the revamped hyperinstrument, known as the Electronic Sitar (*ESitar*). Section 3 describes the building of the robotic Indian drummer, known as *MahaDeviBot*. Section 4 describes experimentation and algorithims toward "intelligent" machine musicianship. Section 5 contains conclusions and future work.

2. THE ELECTRONIC SITAR

The sitar is the prevalent stringed instrument of North Indian classical music traditionally employed to perform ragas. It is distinguished by its gourd resonating chamber, sympathetic strings, and curved frets that allow the incorporation of microtones into melodic phrasing. The first author's initial work on transforming the sitar into a hyperinstrument is described in [17], which serves as a source to gain a more detailed background on traditional sitar performance technique, the sitar's evolution with technology, and initial experimentation and design in building a controller out of a sitar.

The upgraded *ESitar* was designed using new methods and theory obtained from three years of experience of touring and performing. The first step was to find a sitar maker in India to custom design an instrument with modifications to help encase the electronics. One major change to the traditional sitar, was the move to worm-gear tuning pegs for the six main strings. This allows the sitar to remain in tune through all the intense bending during performance, and makes the instrument more accessible to western music students. A second *tumba* (gourd) was also created to encase a speaker to allow for digital sound to resonate through the instrument as well as serve as a monitor for the performer. The bridge, traditionally made of ivory, and then deer bone was upgraded to black ebony wood from Africa, which generates an impressive clear sound and requires less maintenance. The frets themselves were pre-drilled to allow easy installation of the resistor network.

The new *ESitar* made a platform change from the Atmel[1] to the PIC[2] microcontroller, based on the mentoring of the second author. A massive improvement was encasing the microchip, power regulation, sensor conditioning circuits, and midi out device in a box that fits behind the tuning pegs on the sitar itself. This reduces the number of wires, equipment, and complication needed for each performance. This box also has two potentiometers, six momentary buttons, and four push buttons for triggering and setting musical parameters.

The *ESitar* uses a resistor network for fret detection. Military grade resistors at 1% tolerance were used in this new version for more accurate results. Soldering the resistors to the pre-drilled wholes in the frets provided for a more reliable connection that does not have to be re-soldered at every sound check. A force sensing resistor used to obtain thumb pressure proves to be useful in obtaining rhythmic data and pluck direction. A long force sensing resistor which measures force and position was added to the instrument to capture advanced plucking techniques and serve as a digital strum generator. There is a 3-axis accelerometer embedded in the controller box at the top of the neck, to capture ancillary sitar movement, as well as serve as yet another means to control synthesis and audio effect parameters.

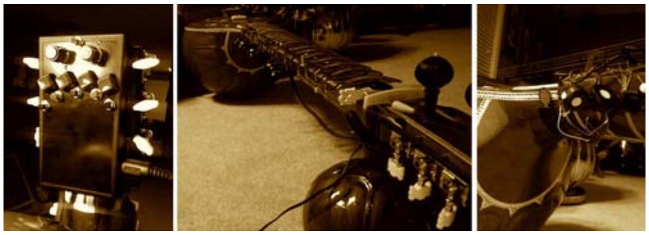

Figure 1 - The upgraded Electronic Sitar Controller.

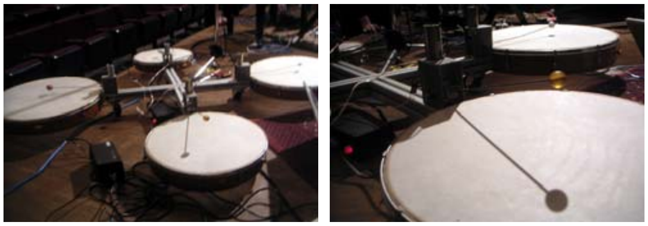

Figure 2 - *MahaDeviBot* Robotic Indian Drummer.

3. ROBOTIC INDIAN DRUMMER

Under the mentorship of the second and fourth author, a robotic Indian drummer, known as *MahaDeviBot* was designed. Modeled after the metaphor of the mythological Hindu Goddess *Saraswati* (Goddess of Art, Music and Knowledge) this robot has multiple arms performing a number of different instruments from India, including frame drums, shakers, bells, and cymbals. The robot is framed using 80/20 T-slotted aluminum[3], for its modularity in design, ease of assembly and disassembly for transportation from venues and laboratories, and for its moderate price. The lower level of the robot has four arms, each playing its own frame drum. Each drum is struck by a solenoid system designed by the second author, commonly found in ModBot [18] projects across New York.

4. MACHINE MUSICIANSHIP

Under the mentorship of the third author, a series of performance scenarios using custom written software were designed to interface the *ESitar* with *MahaDeviBot*. This section will describe three experiments towards "intelligent" machine musicianship: (1) a symbolic MIR-based approach, (2) an audio driven MIR-based approach, (3) and tempo tracking experiments.

4.1 A Symbolic MIR-based Approach

The Music Information Retrieval (MIR) community[4] inspired our initial framework and experimentation for this approach. The goal of this system is to generate a variety of rhythmic accompaniment that evolves over time based on human performance by using sensors to query databases of pre-composed beats. To achieve this, symbolic event databases (shown in Figure 3) for each robotic instrument were filled with rhythmic phrases and variations. During performance, at any given time, queries are generated by sensor data captured from the human performer. As this software is written in ChucK [19], it was easy for the databases to be time and tempo locked to each other to allow for multiple permutations and combinations of rhythm. Figure 3 shows an example of how the system can be mapped. In this case, thumb pressure from the *ESitar* queries what rhythm robotic instrument 1 (*Dha* strokes) will mechanically play on the low frame drum. It is possible to generate a large number of combinations and permutations of rhythms by accessing patterns in each database. This proved to be a successful technique for performances on stage[5].

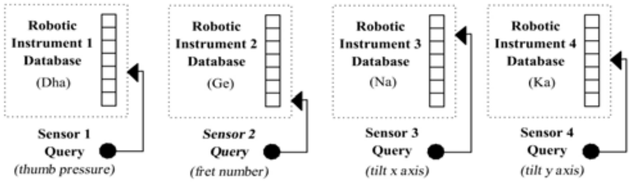

Figure 3 - Symbolic MIR-based approach showing how *ESitar* sensors are used as queries to multiple robotic drum rhythm databases.

[1] http://www.atmel.com/ (January 2007)

[2] http://www.microchip.com/ (January 2007)

[3] http://www.8020.net/ (January 2007)

[4] http://www.ismir.net/ (January 2007)

[5] Videos Available at: http://www.karmetik.com
(Technology → Robotics Department)

One issue to address is how the queries are generated. Methods for finding derivatives and second derivatives for each sensor are employed as queries. Also there are more advanced feature extraction methods, for example obtaining inter onset interval values between peaks of the thumb pressure data. There are many algorithms that can be explored; however the main philosophical question is whether the human should have full control of the machine's performance.

4.2 An Audio Driven MIR-based Approach

A major drawback to the system described in the section above is that rhythms have to be pre-programmed symbolically "by-hand". This motivated this second approach, to have the machine automatically fill its databases by listening to pre-recorded drum beats, music, or even take audio input from a live performer. We chose to write this software in Marsyas [20] because of its strong audio analysis capabilities. This section will first describe the algorithm used to automatically fill these databases, and then describe how sensor data is used to query these databases.

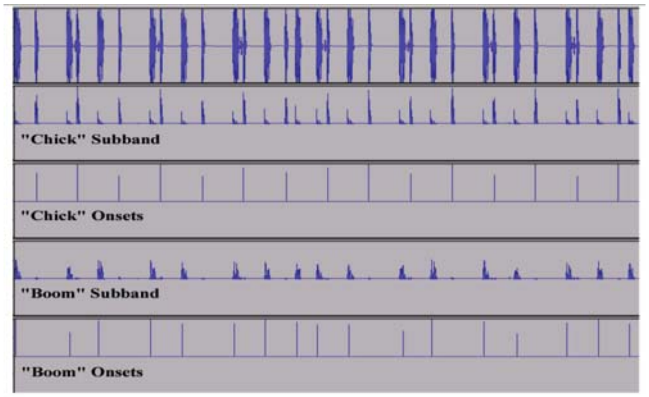

Figure 4 - Audio Signal "Boom-Chick" Decomposition

4.2.1 "Boom-Chick" Audio Analysis

In order to extract symbolic information for the databases we perform what we term "boom-chick" analysis. The idea is to detect the onset of low frequency events typically corresponding to bass drum hits and high frequency events typically corresponding to snare drum hits from the polyphonic audio sources. This is accomplished by onset detection using adaptive thresholding and peak picking on the amplitude envelope of two of the frequency bands.

In order to detect drum sounds two subbands ("boom" and "chick") are utilized. The signal is analyzed in windows of approximately 3 seconds. The frequency range for the low subband is 30Hz to 280Hz. This was determined empirically to give good results over a variety of drum loops and music recordings. A simple bandpass filter implemented in the frequency domain was used to select the "boom" band. The frequency range for the high subband is 2700 Hz- 5500 KHz. The "chick" subband is implemented using a Butterworth filter. It is trivial to design two more filters to create rhythmic events for the remaining drums, however our initial design only has two filters for controlling two robotic drums.

Once the subbands are calculated, they are subsequently processed to find the onset of the "boom" and "chick" sounds. First the envelope of each subband is calculated by using full wave rectification followed by low pass filtering and normalization. Once the envelope is extracted an adaptive peak detection algorithm based on thresholding is utized to find the onset times of the percussive sounds. If the spacing between adjacent peaks is small, then only the highest one will be selected. Figure 4 shows how a drum loop can be decomposed into "boom" and "chick" bands and the corresponding detected onsets. More detailed results on the accuracy of this algorithm can be found in [21] and [22]. Using this software, it is feasible to have the robots play along with any piece of recorded music (preferably rhythmic in nature).

4.2.2 Sensor-Based Retrieval

Now that we have a system to convert audio signals into symbolic data that can be used for robotic performance, the next step is to create a meaningful retrieval method for real-time performance.

The first step is having a human play the ESitar along with each beat stored in the audio database. The sensor data is recorded and aligned with the corresponding symbolic rhythm. For our initial experiments, we chose to use the thumb sensor as our query key, as seen in Figure 5. Before a performance, a database of symbolic rhythms for "Boom" and "Chick" are matched with the corresponding thumb sensor data.

During live performance, a window of thumb sensor data is collected forming a query. This query is then compared with all the pre-recorded thumb sensor data in the database, by doing a simple correlation calculation. The closest match is used to select the corresponding symbolic "boom" and "chick" rhythm for performance by *MahaDeviBot*.

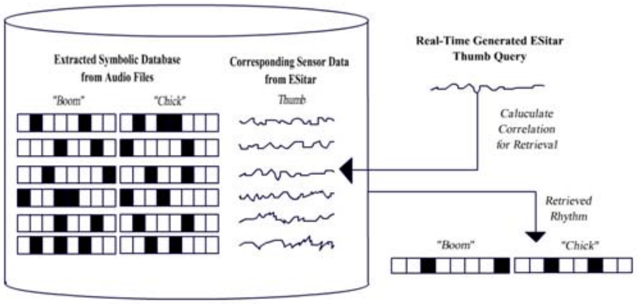
Extracted Symbolic Database from Audio Files
"Boom"
"Chick"
Corresponding Sensor Data from ESitar
Thumb
Real-Time Generated ESitar Thumb Query
Caluculate Correlation for Retrieval
Retrieved Rhythm
"Boom"
"Chick"

Figure 5 - Audio Driven Retrieval Approach.

4.3 Tempo Tracking Experiments

Tempo tracking of the human performer is very crucial for live interactive performance. Humans playing together vary tempo at all times which lets the music breathe. It is easy for the trained musician to follow the tempo of the robot, however, getting the robot to follow the tempo of the human is much more challenging. This section describes our experimentation with comparing tempo extracted from the audio signal and thumb sensor data.

While playing, the performer listened to a constant tempo metronome through headphones. 104 trials were recorded, with each trial lasting 30 seconds. Trials were evenly split into 80, 100, 120, and 140 BPM. The performer would begin each trial by playing a scale at a quarter note tempo, and then a second time at double the tempo. The rest of the trial was an improvised session in tempo with the metronome. Audio was captured using a pick-up on the *ESitar* and recorded at a sample rate of 44100 Hz. The corresponding thumb sensor data was recorded at a sample rate of 86 Hz.

Onset detection algorithms were applied to the audio and sensor signals to gather information about when the strings were strummed. To begin, the RMS energy of the audio signal are calculated while the values of the thumb sensor are used directly. A peak-picking algorithm is applied to the derivatives of each signal to find onset locations. Real-time tempo tracking is performed using a probabilistic Particle Filter. The algorithm tests various hypotheses of the output of a switching Kalman filter against noisy onset measurements providing an optimal estimate of the beat period and beat. Analysis of accuracy of various methods of achieving this goal is presented in Figure 6. For each signal (audio and thumb) we extract onsets that are subsequently processed by Kalman filtering [23] for tempo tracking [24]. As can be seen in the table, late fusion (combining tempo estimations of both audio and thumb signals) is shown to be superior to using each signal individually. This helps justify why we need a hyperinstrument to accomplish our goal.

Ground Truth	**80**	**100**	**120**	**140**
Audio	*46%*	*85%*	*86%*	*80%*
Thumb	*35%*	*62%*	*75%*	*65%*
Late Fusion	*57%*	*88%*	*90%*	*80%*

Figure 6 - Result of Audio, Thumb, Late Fusion Tempo Extraction.

5. CONCLUSION & FUTURE WORK

This paper presented an updated version of a controller introduced in 2004, modified using the influence and knowledge gained from the NIME community over the years. It also presents the beginning a robotic Indian drummer known as *MahaDeviBot*. Most importantly, it presents initial experimentation and algorithms to interface a human performer and robot.

There are many directions for future work. The *MahaDeviBot* is in initial stages of adding nine more arms, making a total of thirteen. Also, more advanced functions for robotic expression like striking in different locations on the drum are planned using RC stepper motors. It is also interesting to explore rhythmic similarity and extract queries based on rhythm. We also plan to look at pitch recognition from the ESitar and how it can be interfaced into the system. Also, we are exploring ways to create the databases of rhythm in real-time during the performance. The overall strategy to accomplish this is to keep booking performances, improving one or two elements of the system in the laboratory, and testing it live on the performance stage.

6. ACKNOWLEDGMENTS

Thanks to Ustad Siraj Khan and Rakesh Kumar Parihast for their training in Indian Classical theory. Special thanks to Shahid the sitar maker in Miraj, India. Thanks to Rodney Katz for help with cutting aluminum pieces. Thanks to Perry Cook, Ge Wang for writing ChucK, and their constant support. Thanks to Peter Driessen, Andrew Schloss, and the MISTIC team at University of Victoria for their teamwork and support.

7. REFERENCES

[1] Trimpin, *SoundSculptures: Five Examples*. Munich MGM MediaGruppe Munchen, 2000.

[2] J. Solis, M. Bergamasco, S. Isoda, K. Chida, and A. Takanishi, "Learning to Play the Flute with an Anthropomorphic Robot," in *International Computer Music Conference (ICMC)*, Miami, Florida, 2004.

[3] S. Jorda, "Afasia: the Ultimate Homeric One-Man-Multimedia-Band," in *International Conference on New Interfaces for Musical Expression (NIME)*, Dublin, Ireland, 2002.

[4] C. MacMurtie, "Amorphic Robot Works."

[5] N. A. Baginsky, "The Three Sirens: A Self Learning Robotic Rock Band."

[6] G. W. Raes, "Automations by Godfried-Willem Raes."

[7] A. Kapur, "A History of Robotic Musical Instruments," in *ICMC*, Barcelona, Spain, 2005.

[8] T. Machover and J. Chung, "Hyperinstruments: Musically Intelligent and Interactive Performance and Creativity Systems," in *ICMC*, 1989, pp. 186-190.

[9] D. Trueman and P. R. Cook, "BoSSA: The Deconstructed Violin Reconstructed," in *ICMC*, Beijing, China, 1999.

[10] R. Rowe, *Machine Musicianship*. Cambridge, MA: MIT Press, 2004.

[11] R. B. Dannenberg, "An On-line Algorithm for Real-Time Accompaniment," in *ICMC*, Paris, France, 1984, pp. 193-198.

[12] G. Lewis, "Too Many Notes: Computers, Complexity and Culture in Voyager," *Leonardo Music Journal*, vol. 10, pp. 33-39, 2000.

[13] F. Pachet, "The Continuator: Musical interaction with Style " in *ICMC*, Goteborg, Sweden, 2002.

[14] E. Singer, K. Larke, and D. Bianciardi, "LEMUR GuitarBot: MIDI Robotic String Instrument," in *NIME*, Montreal, Canada, 2003.

[15] G. Weinberg, S. Driscoll, and M. Parry, "Haile - A Preceptual Robotic Percussionist," in *ICMC*, Barcelona, Spain, 2005.

[16] G. Weinberg, S. Driscoll, and T. Thatcher, "Jam'aa - A Middle Eastern Percussion Ensemble for Human and Robotic Players," in *ICMC*, New Orleans, 2006, pp. 464-467.

[17] A. Kapur, A. Lazier, P. Davidson, R. S. Wilson, and P. R. Cook, "The Electronic Sitar Controller," in *NIME*, Hamamatsu, Japan, 2004.

[18] E. Singer, J. Feddersen, C. Redmon, and B. Bowen, "LEMUR's Musical Robots," in *NIME*, Hamamatsu, Japan, 2004.

[19] G. Wang and P. R. Cook, "ChucK: A Concurrent, On-the-fly Audio Programming Language," in *ICMC*, Singapore, 2003.

[20] G. Tzanetakis and P. R. Cook, "Marsyas: a Framework for Audio Analysis," *Organized Sound*, vol. 4, 2000.

[21] A. Kapur, R. I. McWalter, and G. Tzanetakis, "New Interfaces for Rhythm Based Information Retrieval," in *International Conference on Music Information Retrieval*, London, UK, 2005.

[22] G. Tzanetakis, A. Kapur, and R. I. McWalter, "Subband-based Drum Transcription for Audio Signals," in *IEEE International Workshop on Multimedia Signal Processing*, Shanghai, China, 2005.

[23] T. Cemgil, "Bayesian Music Transcription." vol. Ph.D. Netherlands: Radboud University of Nijmegen, 2004.

[24] T. v. Kasteren, "Realtime Tempo Tracking using Kalman Filtering," in *Computer Science*. vol. Masters Amsterdam: University of Amsterdam, 2006.