# A Comparative Study on Wearable Sensors for Signal Processing on the North Indian Tabla

Manjinder S. Benning[1,3] Ajay Kapur[1,3] Bernie Till[2] George Tzanetakis[1] Peter Driessen[1]
University of Victoria Music Intelligence and Sound Technology Interdisciplinary Centre (MISTIC)[1]
University of Victoria Assistive Technology Team (UVATT)[2]
KarmetiK Technology (A Division of KarmetiK LLC)[3]
Victoria, British Columbia, Canada

***Abstract*—This paper describes experimentation using a variety of sensor techniques to capture body gestures and train a student performing the North Indian hand drums known as the Tabla. A comparative study of motion capture systems, wearable accelerometer units, and wireless inertial sensor packages, is described. Each acquisition method has it advantages and disadvantages which are explored through trial and error. The paper describes a number of applications using real-time signal processing techniques for analysis, performance, performer posture detection and machine perception of human interaction.**

## I. Introduction

The motion of the human body is a rich source of information, containing intricacies of musical performance which can aid in obtaining knowledge about intention and emotion of human interaction with an instrument. This paper explores a variety of techniques for obtaining data from a performing artist by placing sensors on the human body. The sensor data is used for a variety of applications including sonification of body gestures for analysis, real-time control of synthesis and audio effect parameters, posture correction of a training musician and real-time tempo tracking systems.

In particular, this paper focuses on one musical instrument, the North Indian hand drum known as the Tabla. However algorithms are modular, and can work on any traditional instrument. There has been work on obtaining musical information from a tabla performer using the custom built drums known as the ETabla [1]. Another similar paradigm is that of the hyperinstrument [12, 13] where an acoustic instrument is augmented with sensors. In our approach, any performer can wear a sensor while keeping the acoustic instrument unmodified, allowing a more accessible and flexible systems.

Section II will describe experiments using a motion capture system. Section III will describe an evolution to using a wearable sensor package to obtain acceleration data. Section IV will describe yet another evolution to a multiple orientation wireless sensor package system that obtains human body posture. Experiments with tabla performers are included throughout the paper.

## II. Sonification with Motion Capture

This section describes experiments using a motion capture system to help begin understanding the intricacies of human body motion during musical performance. Inspired by the work of Marcelo Wanderly at McGill University in Montreal, Canada, who uses motion capture systems to study three main factors that influence performance: (1) The instruments constraints on the body, (2) the characteristics of the performance (e.g. rhythm, articulation, tempo, etc.) and, (3) the interpretive momentary choices of the performer [2, 3].

Our first goal was to build the necessary infrastructure to study the use of sonification for understanding human motion in a musical context. In order to achieve this, VICON[1], a commercial vision-based motion capturing system was interfaced with various sound producing languages and frameworks. Sonification of human motion can yield results that are not observable by vision alone. Perception of periodicity, regularity, and speed of motion are few of the attributes that are easier to observe with the aid of sound to aid design decisions of real-time software applications.

Although the proposed infrastructure has been applied to many areas of research[5], the goals relating to this paper include studying how a musician's posture and gestural movements during performance affect the sound produced as well as the emotional content [6] of the performer.

### A. VICON Motion Capture System

The VICON Motion Capture System is designed to track human or other movement in a room-size space. Spheres covered with reflective tape, known as markers, are placed on visual reference points on different parts of the human body. The VICON system consists of 6 cameras and is designed to track and reconstruct these markers in 3-dimensional space. When a marker is seen by one of the cameras, it will appear in the camera's view as a series of highly illuminated pixels in comparison to the background. During capture the coordinates of all the markers in each camera's view are stored in a data-

[1] http://www.vicon.com (Available January 2005)

station. The VICON system then links the correct positions of each marker together to form continuous trajectories, which represent the paths that each marker has taken throughout the capture and thus how the subject has moved over time. At least three of the cameras must view a marker for the point to be captured. Continuous signals are obtained through interpolation [7]. The VICON system measured the trajectories of each subject's movement in 3D space at a sampling rate of 120 Hz, however newer systems have much higher sample rates (1-2 kHz) for more precise data.

### B. *Experimental Procedure*

After motion capture trials are run using the VICON system, all data is labeled and interpolation algorithms are run to obtain continuous streams of marker positions. Next, each trial is exported to a text file. The first line of the text file contains the label names of the markers, delineated by a comma. Each line is time stamped and represents the x-, y-, z-, coordinates of all the markers for that particular time instance.

Tabla performance recordings were of traditional *Tin Taal Theka* excerpts of 16-beat cycles. As shown in Figure 1, a full model of the right hand was captured using a custom built VICON plugin to capture 28 marker points. Once the data is collected, the files are imported into the desired synthesis synthesis languages: STK[8], Marsyas[9], and ChucK[10].

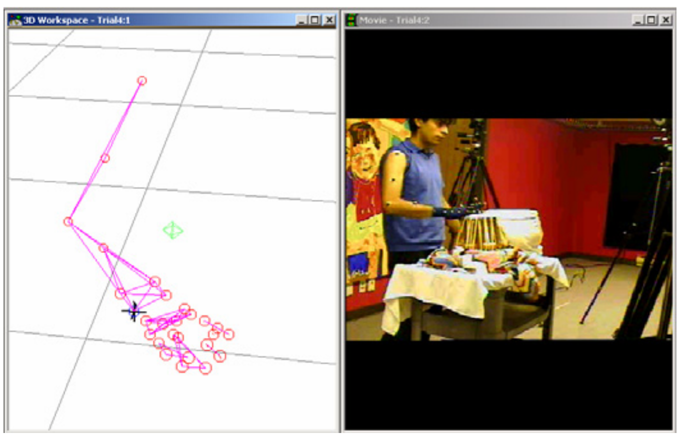


Figure 1 - Screenshot of data capturing process for tabla performance.

### C. *Discussion*

First, we are interested in finding which markers contain musical information. This is tested using STK's physical models of the instrument, in order to try and reproduce a performance, using the marker's data to control parameters of appropriate physical models. Our goal is to find how few markers can be used in order to reconstruct a musical phrase. Another interesting question is to observe interchanging traditional mappings (e.g. map plucking hand to bowing, and bowing hand to plucking) to obtain new types of sound.

Another area of interest is to observe ancillary gestures during performance (e.g. how the head moves during a performance). Specifically, the following questions are asked: When a performer plays the same composition, do the ancillary body gestures move in the same way? What is the minimum number of markers that need to be the same in order for the same performance to be played? What type of information do the ancillary markers obtain? Answering these questions using the proposed framework allows observations of subtle differences in movements that are difficult to see using only visual feedback. The challenge with the tabla is the precise timing of fingers. Thus we use a detailed model of the hand, as described above, in order to preserve the performance.

## III. Real-time Modulation and Tempo Tracking with the KiOm

The VICON Motion Capture System provides an immense amount of data for analysis and research. However, the system used is not real-time (although real-time versions of hardware/software commercially exist). VICON system and other motion capture systems are very expensive, and cumbersome/impossible to move on stage for performance. Also, the markers which are stuck on the musicians tend to fall off and lighting conditions must meet requirements for ideal capture. These drawbacks have influenced the invention of the KiOm [11] wearable sensor. This section describes the use of wearable sensor technology to control parameters of audio effects for real-time musical signal processing and real-time tempo tracking of a tabla performance.

Traditional instrument performance techniques are preserved while the system modifies the resulting sound based upon the movements of the performer. Gesture data from a performing artist is captured using a three-axis accelerometer packages that is converted to MIDI (Musical Instrument Digital Interface) messages using microcontroller technology. Traditionally, sensors are used to drive synthesis algorithms directly, completely separating the sound source from the gesture. Our paradigm, and the key novelty of this work, is to keep traditional instrument performance technique, modifying the amplified acoustic signal with sensor data controlling a number of audio effect parameters.

As well as musical signal processing on the audio signal, onsets of the KiOm data along with onsets from the RMS of the audio signal are used as input into a Kalman Filter based real-time tempo tracking system [16]. For this experiment a KiOm was attached to the top of the right hand wrist to capture the movements of a tabla performer to aid in machine perception of tempo.

### A. *The KiOm*

The *KiOm* is built using a Kionix KXM52-1050[2] three-axis accelerometer. The three streams of analog gesture data from

[2] http://www.kionix.com/ (February 2005)

the sensor is read by the internal ADC of the Microchip PIC 18F2320[3]. These streams are converted to MIDI messages for use with most musical hardware/synthesizers.

### *B. Discussion of Real-Time Effect Modulation*

The wearable prototype sensor was first placed on the hands of a tabla (North Indian drum) player during performance. The drummer was told to play with traditional technique. Because of the rhythmic nature and movement of the drummer's hands during the performance, using the gesture-captured data to effect the sounds of the drums was successful. Our favorite algorithms were controlling parameters of ring modulation and reverb. Similar results were obtained by placing the sensors on the feet of the drummer while playing bass drum.

### *C. Tempo Tracking Experimental Procedure*

For our experiments we recorded a data set of a performer playing the tabla with a KiOm on the right hand. Audio files were captured at a sampling rate of 44100 Hz. 3 axes of acceleration data in MIDI format were also recorded. While playing, the performer listened to a constant tempo metronome through headphones. 52 trials were recorded, with each trial lasting 30 seconds. Trials were evenly split into 80, 100, 120, and 140 BPM, using the metronome connected to the headphones. The performer would begin each trial by playing a *Tin Taal Theka* 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.

Onset detection algorithms were applied to the audio and sensor signals to gather periodicity information of the music performance. The RMS energy of the audio signal and the magnitude of the KiOm 3-axes accelerations are calculated. A peak-picking algorithm is applied to the derivatives of each signal to find onset locations. An adaptive peak-picking algorithm is applied on the KiOm data to compensate for the large variability in wrist movement during performance. In order to detect this underlying tempo period we utilize a real-time Kalman filtering algorithm for each sequence of onsets transmitted as MIDI signals.

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 [16]. Noisy onset measurements, extracted from the various sensor streams, are used as input to a real-time implementation of the tempo tracking algorithm [17]. In order to model the onset sequence we use a linear dynamical system as proposed by Cemgil, see [16] for a detailed explanation of the tempo tracker. In the following section the accuracy of the two estimated beat period streams is evaluated. In addition we show that late fusion of the streams can significantly improve tempo detection accuracy.

### *D. Results*

Table I shows the percentages of frames for which the tempo was correctly estimated. Tempo estimates are generated at 86Hz resulting in approximately 2600 estimates/30 second clip in the dataset. From the percentages of Table I, we can conclude that when using a single acquisition method, the Audio and KiOm obtained the similar results at slower tempos, and the audio signal was best for faster tempos.

TABLE I. TEMPO TRACKING RESULTS

| **Signal** | **Tempo (BPM)** | | | |
|---|---|---|---|---|
| | ***80*** | ***100*** | ***120*** | ***140*** |
| Audio | 67% | 86% | 79% | 72% |
| KiOm | 65% | 81% | 68% | 60% |
| LATE FUSION: | | | | |
| Audio/ KiOm | 91% | 93% | 85% | 80% |

When looking carefully through the detected onsets from the two types of acquisition methods, we observed that they exhibit outliers and discontinuities at times. To address this problem we utilize a late fusion approach where we consider each acquisition method in turn for discontinuities. If a discontinuity is found, we consider the next acquisition method to make a final decision. By fusing the acquisition methods together, we are able to get more accurate results. At all BPM's the late fusion of the Audio and KiOm data streams generate stronger tempo tracking results than either stream individually. The most improvement can be seen at 80 BPM's.

## IV. POSTURE DETECTION WITH THE WISP

The Wireless Inertial Sensor Package (WISP) [14] was designed by the third author specifically for the task of capturing human body movements. A WISP was attached to the right elbow of a tabla student to help them achieve correct right arm posture. Also, Multiple WISP's are being experimented with to obtain posture information of the head, neck, back, and left arm.

### *A. The WISP*

The WISP is a highly integrated IMU with on-board DSP and radio communication resources. It consists of a triaxial differential capacitance accelerometer, a triaxial magnetoresistive bridge magnetometer, a pair of biaxial vibrating mass coriolis-type rate gyros, and a NTC thermistor. This permits temperature-compensated measurements of linear acceleration, orientation, and angular velocity. The first generation prototype of WISP, shown in Figure 2 next to a Canadian two-dollar coin, uses a 900 MHz transceiver with a 50Kb/s data rate. With a volume of less than $13cm^3$ and a mass

[3] http://www.microchip.com/ (February 2005)

of less than 23g, including battery, the unit is about the size of a somewhat large wristwatch. The WISP can operate for over 17 hours on a single 3.6V rechargeable Lithium cell, which accounts for over 50% of the volume and over 75% of the mass of the unit. As can be seen in Figure 2, the small size and flat form-factor make it ideal for unobtrusive, live and on-stage, real-time motion-capture.

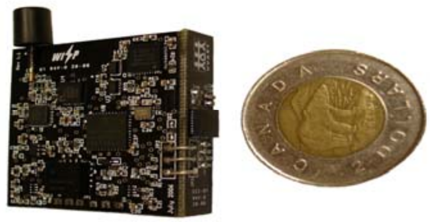

Figure 2 - Wireless Inertial Sensor Package (WISP).

### *B. Experimental Procedure*

Orientation data of the WISP, communicated via the OSC protocol [15], was used to give feedback to the tabla student to help them maintain correct right arm posture during playing. In the beginner stages of tabla, body posture, particularly orientation of the right arm is critical to the correct development of a student. Boundaries were imposed on the 3 axis of orientation given by the WISP. When any of the boundaries were exceeded by the student an axis specific sound would alert the student of incorrect posture.

Initial work also includes an animated human body model instrumented by 7 WISPs attached to a student's upper body at key skeletal articulations. Since each WISP is small and wireless this sort of whole-body gesture analysis is easy to implement and non-invasive for the student. A multi WISP system will provide a much more detailed analysis of posture enabling the system to teach a student to maintain correct posture of the spine, neck, head, and left arm as well as the right arm. And of course this system can be extended beyond the tabla into any sort of posture critical applications.

TABLE II. COMPARISON OF AQUISITION METHODS

| Device | Advantages | Disadvantages |
|---|---|---|
| *Motion Capture* | * Rich Amounts of Data<br>* No Electonics on Body | * Expensive<br>* Cumbersome for Stage<br>* Markers fall off<br>* Reflection Errors |
| *KiOm* | * Cheap<br>* Accesible<br>* MIDI Device Compatible | * Electronics on Body<br>* Wired<br>* Accleration data only |
| *WISP* | * Orientation data<br>* Wireless<br>* OSC Device Compatible | * Moderately Expensive<br>* Electronics on Body |

## V. CONCLUSION

This paper showed comparisons between three different methods of capturing gestures of a tabla performer. In each case the captured data was used in different applications including sonification, audio effect control, real-time tempo tracking and performer/student posture training. Refer to table II for a summary of advantages and disadvantages of each method. Our experiments show that gesture data of a musician is paramount in the evolution of musical signal processing.

## REFERENCES


[1] A. Kapur, G. Essl, P. Davidson, and P. R. Cook, "The Electronic Tabla Controller," *Journal of New Music Research,* vol. 32, pp. 351-360, 2003.

[2] M. M. Wanderley, "Non-Obvious Performer Gestures in Instrumental Music," in *Gestrure-Based Communication in Human-Computer Interaction*, 1999, pp. 37-48.

[3] M. M. Wanderley, "Quantitative Analysis of Non-obvious Performer Gestures," in *Gesture and Sign Language in Human-Computer Interaction*, 2002, pp. 241-253.

[4] M. Peinado, B. Herbelin, M. M. Wanderley, B. L. Callennec, D. Thalmann, and D. Meziatil, "Towards Configuarable Motion Capture with Polorized Inverse Kinematics " *Sensor* 2004.

[5] A. Kapur, G. Tzanetakis, N. Virji-Babul, G. Wang, and P. R. Cook, "A Framework for Sonification of Vicon Motion Capture Data," in *International Conference on Digital Audio Effects*, Madrid, Spain, 2005.

[6] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. F. Driessen, "Gesture-Based Affective Computing on Motion Capture Data," in *International Conference on Affective Computing and Intelligent Interaction*, Beijing, China, 2005, pp. 1-7.

[7] A. Woolard, *Manual*. Tustin, CA: VICON Motion Systems, 1999.

[8] P. R. Cook and G. Scavone, "The Synthesis Toolkit (STK)," in *International Computer Music Conference (ICMC)*, Beijing, China, 1999.

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

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

[11] A. Kapur, E. L. Yang, A. R. Tindale, and P. F. Driessen, "Wearable Sensors for Real-Time Musical Signal Processing," in *IEEE Pacific Rim Conference*, Victoria, Canada, 2005.

[12] A. Kapur, A. Lazier, P. Davidson, R. S. Wilson, and P. R. Cook, "The Electronic Sitar Controller," in *International Conference on New Interfaces for Musical Expression (NIME)*, Hamamatsu, Japan, 2004.

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

[14] B. C. Till, M. S. Benning, and N. Livingston, "Wireless Inertial Sensor Package (WISP)," in *NIME*, New York City, 2007.

[15] M. Wright, A. Freed, and A. Momeni, "OpenSound Control: State of the Art 2003," in *NIME*, Montreal, Canada, 2003.

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

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