NEW MUSIC INTERFACES FOR RHYTHM-BASED RETRIEVAL

Ajay Kapur
University of Victoria
3800 Finnerty Rd.
Victoria BC, Canada
ajay@ece.uvic.ca

Richard I. McWalter
University of Victoria
3800 Finnerty Rd.
Victoria BC, Canada
rmcwalte@ece.uvic.ca

George Tzanetakis
University of Victoria
3800 Finnerty Rd.
Victoria BC, Canada
gtzan@cs.uvic.ca

ABSTRACT

In the majority of existing work in music information retrieval (MIR) the user interacts with the system using standard desktop components such as the keyboard, mouse or sometimes microphone input. It is our belief that moving away from the desktop to more physically tangible ways of interacting can lead to novel ways of thinking about MIR. In this paper, we report on our work in utilizing new non-standard interfaces for MIR purposes. One of the most important but frequently neglected ways of characterizing and retrieving music is through rhythmic information. We concentrate on rhythmic information both as user input and as means for retrieval. Algorithms and experiments for rhythm-based information retrieval of music, drum loops and indian tabla thekas are described. This work targets expert users such as DJs and musicians which tend to be more curious about new technologies and therefore can serve as catalysts for accelerating the adoption of MIR techniques. In addition, we describe how the proposed rhythm-based interfaces can assist in the annotation and preservation of perfomance practice.

Keywords: user interfaces, rhythm analysis, controllers, live performance

1 INTRODUCTION

Musical instruments are fascinating artifacts. For thousands of years, humans have used all sorts of different materials including wood, horse hairs, animal hides, and bones to manufacture a wide variety of musical instruments played in many different ways. The strong coupling of the musicians' gestures with their instruments was taken for granted for most of music history until the invention of recording technology which made it possible to listen to music without the presence of performers and their instruments.

Music Information Retrieval (MIR) has the potential of revolutionizing the way music is produced, archived and consumed. However, most of existing MIR systems are prototype systems developed in academia or research labs that have not yet been empbraced by the public. One of the possible reasons is that so far most of existing systems have focused on two types of users: 1) users that have knowledge of western common music notation such as musicologists and music librarians or 2) "average users" who are not necessarily musically-trained. This bias is also reflected in the selection of problems typically involving collections of either western classical music or popular music.

In this paper, we explore the use of non-standard interaction devices for MIR. These devices are inspired by existing musical instruments and are used for both retrieval and browsing. They attempt to mimic and possibly leverage the tangible interaction of performers with their instruments. Therefore the target users are musicians and DJs which in our experience tend to be more curious about new technologies and therefore can serve as catalysts for accelerating the adoption of MIR techniques.

Although the main ideas we propose are applicable to any type of instrument-inspired interaction and music retrieval tasks the focus of this paper is the use of rhythmic information both as user input and as means for retrieval. Rhythm is fundamental in understanding music of any type and provided us with a common thread behind this work. The described interfaces are inspired by existing musical instruments and have their origins in computer music performance. They utilize a variety of sensors to extract information from the user. This information is subsequently used for browsing and retrieval purposes. In addition these sensor-enhanced instruments can be used to archive performance-related information which is typically lost in audio and symbolic representations. The integration of the interfaces with MIR algorithms into a prototype system is described and experimental results using collections of drum loops, tabla thekas and music pieces of different genres are presented. Some reprsentative scenarios are provided to illustrate how the proposed interfaces can be used. It is our hope that these interfaces will extend MIR beyond the standard desktop/keyboard/mouse interaction into new contexts such as practicing musical instruments, live performance and the dance hall.

2 RELATED WORK

There are two main areas of related work: 1) non-standard tangible user interfaces for information retrieval and 2) rhythm analysis and retrieval systems.

Using sensor-based user interfaces for information retrieval is a new and emerging field of study. The bricks project (Fitzmaurice et al., 1995) at MIT is an early example of a graspable user interface being used to control virtual objects, such as objects in a drawing program.

One of the most inspiring interfaces for our work is *musicBottles*, a tangible interface designed by Hiroshi Ishii and his team at the MIT Media Lab. In this work, bottles can be opened and closed to explore a music database of classical, jazz, and techno music (Ishii, 2004). Ishii elegantly describes in his paper his mother's expertise in "everyday interaction with her familiar physical environment - opening a bottle of soy sauce in the kitchen." His team thus built a system that took advantage of this expertise, so that his mother could open a bottle and hear birds singing to know that tomorrow would be a sunny, beautiful day, rather than having to use a mouse, keyboard to check the system online. This work combines a tangible interface with ideas from information retrieval. Similarly in our work we try to use existing interaction metaphors for music information retrieval tasks.

The use of sensors to gather gestural data from a musician has been used as an aid in the creation of real-time computer music performance. Examples of such systems include: the Hypercello (Machover, 1992), and the digitized japanese drum Aobachi (Young and Fujinaga, 2004).

Also there has been some initial research on using interfaces with music information retrieval for live performance on stage. AudioPad (Patten et al., 2002) is an interface designed at MIT Media Lab, which combines the expressive character of multidimensional tracking with the modularity of a knob-based interface. This is accomplished by using embedded LC tags inside a puck-like interface which is tracked in two dimensions on a tabletop. It is used to control parameters of audio playback, acting as a new interface for the modern disk jockey. Block Jam (Newton-Dunn et al., 2003) is an interface designed by Sony Research which controls audio playback with the use of 25 blocks. Each block has a visual display and a button-like input for event driven control to control functionality. Sensors within the blocks allow for gesture-based manipulation of the audio files. Researchers from Sony CSL Paris proposed SongSampler (Aucouturier et al., 2004), a system which samples a song and then uses a MIDI instrument to perform the samples from the original sound file.

There has also been some work in using rhythmic information for MIR. The use of *BeatBoxing*, the art of vocal percussion, as a query mechanism for music information retrieval was proposed by Kapur et al. (2004a). A system which classified and automatically identified individual beat boxing sounds, mapping them to corresponding drum samples was developed. A similar concept was proposed by Nakano et al. (2004) in which the team created a system for voice percussion recognition for drum pattern retrieval. Their approach used onomatopoeia as the internal representation of drum sounds which allowed for a larger variation of vocal input with an impressive identification rate. Gillet and Richard (2003) explore the use of the voice as a query mechanism in the different context of Indian tabla music. A system for *Query-by-Rhythm* was introduced by Chen and Chen (1998). Rhythm is stored as strings turning song retrieval into a string matching problem. The authors propose an L-tree data structure for efficient matching. The similarity of rhythmic patterns using a dynamic programming approach is explored by Paulus and Klapuri (2002). A system for the automatic description of drum sounds using a template adaption method is Yoshii et al. (2004).

3 INTERFACES

The musical interfaces described in this paper are tangible devices used to communicate musical information through gestural interaction. They enable a more rich and musical interaction than the standard keyboard and mouse. These interfaces use modern sensor technology such as force sensing resistors, accelerometers, infrared detectors, and piezoelectric sensors to measure various aspects of the human-instrument interaction. Data is collected by an onboard microprocessor which converts the sensor data into a digital protocol for communicating with the computer used for MIR. Currently, the MIDI message protocol is utilized.

3.1 Mercurial STC1000

The Mercurial STC1000 (shown in Figure 1 (A)) [1] uses a network of fiberoptic sensors to detect pressure as position on a two-dimensional plane. It has been designed by the Mercurial Innovations Group. This device is a singe touch controller that directly outputs MIDI messages. The mapping to MIDI can be controlled by the user. The active pad area is 125mm X 100mm (5 X 4 inches).

3.2 Radio Drum

The Radio Drum/Baton, shown in Figure 1 (B), is one of the oldest electronic music controllers (Mathews and Schloss, 1989). Built by Bob Boie and improved by Max Mathews, it has undergone a great deal of improvement in accuracy of tracking, while the user interaction has remained relatively constant. The drum generates 6 separate analog signals that represent the current x, y, z position of each stick. The radio tracking is based on measuring the electrical capacitance between the coil at the end of each stick and the array of receiving antennas on the drum (one for each corner). The analog signals are converted to MIDI messages by a microprocessor. The sensing surface measures approximately 375mm X 600mm (15 X 24 inches).

3.3 ETabla

The traditional tabla are a pair of hand drums used to accompany North Indian vocal and instrumental music. The Electronic Tabla Controller (ETabla) (Kapur et al.,

[1] `http://www.thinkmig.com/stc1000.html`

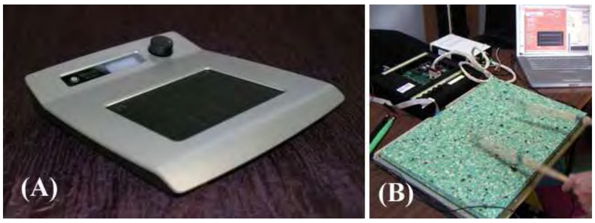

Figure 1: (A) Mercurial STC 1000 and (B) Radio Drum

Figure 2: (A) ETabla (B) ESitar (C) EDholak

2004b) (shown in Figure 2 (A)) is a custom built controller that uses force sensing resistors to detect different strokes, strike velocity and position. Though an acoustically quiet instrument, it adheres to traditional technique and form. The ETabla is designed to allow performers to leverage their skill to control digital information.

3.4 EDholak

The Electronic Dholak Controller (EDholak) (Kapur et al., 2004b) is another custom built controller using force sensing resistors and piezoelectric sensors to capture rhythmic gestures. The Dholak is an Indian folk drum performed by two players. Inspired by the collaborative nature of the traditional drum, the EDholak (shown in Figure 2 (C)) is a two player electronic drum, where one musician provides the rhythmic impulses while the other provides the tempo and controls the sound production of the instrument using a sensor-enhanced spoon. This interface, in addition to sending sensor data, produces actual sound in the same way as it's traditional counterpart.

3.5 ESitar

The Electronic Sitar Controller (ESitar) (Kapur et al., 2004b) is a hyperinstrument designed using a variety of sensor techniques. The Sitar is a 19-stringed, pumpkin shelled, traditional North Indian instrument. The ESitar (shown in Figure 2 (B)) modifies the acoustic sound of the performance using the gesture data deduced from the sensors, as well as serving as a real-time transcription tool that can be used as a pedagogical device. The ESitar obtains rhythmic information from the performer by a force sensing resistor placed under the right thumb, deducing stroke direction and frequency.

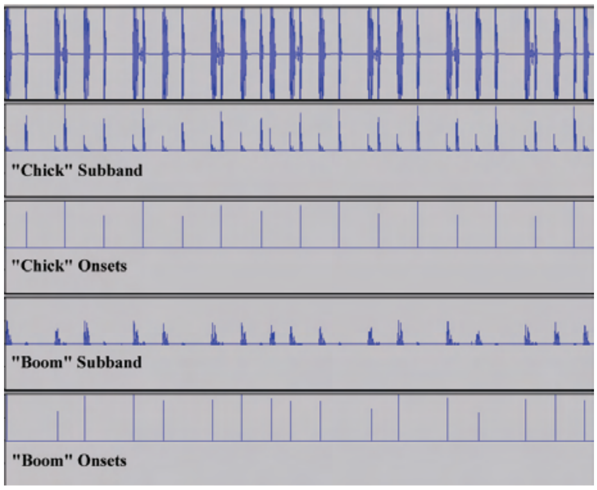

Figure 3: "Boom-chick" onset detection

4 RHYTHM ANALYSIS AND RETRIEVAL

The proposed interfaces generate essentially symbolic data. However our goal is to utilize this information to browse and retrieve from collections of audio signals. Therefore a rhythmic analysis front-end is used to convert audio signals to more structured symbolic representations. This front-end is based on decomposing the signal into different frequency bands using a Discrete Wavelet Transform (DWT) similarly to the method described by Tzanetakis and Cook (2002) for the calculation of *Beat Histograms*. The envelope of each band is then calculated by using full wave rectification, low pass filtering and normalization. In order to detect tempo and beat strength the *Beat Histogram* approach is utilized.

In order to extract more detailed information 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. This is accomplished by onset detection using adaptive thresholding and peak picking on the amplitude envelope of two of the frequency bands of the wavelet transform (approximately 300Hz-600Hz for the "boom" and 2.7KHz-5.5KHz for the "chick"). Figure 3 shows how a drum loop can be decomposed into "boom" and "chick" bands and corresponding detected onsets. Even though more sophisticated algorithms for tempo extraction and drum pattern detection, such as the ones mentioned in the related work section, have been proposed, the above approach worked quite well and provide us with the necessary infrastructure for experimenting with the new interfaces.

The onset sequences of "boom-chick" events can be converted into a string representation for retrieval purposes. Once onset times are found, the follwoing two representations are created:

```
Chick array    --C---C-----C-C----
Boom array:    B---B-B---B---B---B
```

The next step is to combine the two representations into one. If a Bass and Snare occur at the same time, then T represents B+C:

```
Combined Array: B-C-BCB---T-C-B---B
```

This composite representation array is then used to create a string combining the type of onset with durations between each event. There are six types of transitions that are labeled with durations: BC, CB, TC, CT, BT. Durations are relative (similar to common music notation). They are calculated by picking the most common inter-onset interval (IOI) using clustering and expressing all other IOI's after quantization as ratios to the most common one. Typically the most common IOI corresponds to eighth notes or quarter notes. This representation is invariant to tempo. The quantized IOIs form essentially a dictionary of possible rhythmic durations (an hierarchy of 5-6 durations is typically adaquate). In order to represent the "boom-chick" events these durations are combined with the 6 possible transitions to form an alphabet. For example a full beat string can be represented as:

```
{BC2,CB2,BC1,CB1,BT4,TC2,CB2,BB4}
```

In this string representation each triplet essentially corresponds to one character in the alphabet. These strings can then be compared using standard approximate string matching algorithms such as Dynamic Programming. Although straightforward, this approach works quite well for music with strong beats which is the focus of this work.

5 SCENARIOS

In this section, we illustrate using scenarios some of the ways the proposed interfaces can be used for MIR. These scenarios have been implemented as proof-of-concept prototypes and are representative of what is possible using our system. They also demonstrate the interplay between annotation, retrieval and browsing that is made possible by non-standard MIR interfaces. Initial reception by musicians and DJs of our prototypes has been encouraging.

5.1 Tapping Tempo-based Retrieval

One of the most basic reactions to music is tapping. Any of the proposed drum interfaces can be used to generate a tempo-based query. Even in this simple fundamental interaction, being able to tap the rhythm using a stick or a finger on a surface is preferable to clicking a mouse button or keyboard key. The query tempo can directly be measured and compared to a database of tempo-annotated music pieces or drum loops. The annotation can be performed manually or automatically using audio analysis algorithms. Moreover, tempo annotation can also be done using the same process as the query specification. Tempo-based retrieval is also useful for the DJ, saving time by not having to thumb through boxes of vinyl, or scroll through hundreds of mp3s for a song that is at a particular tempo. Tapping the tempo into an interface is more convenient because the DJ can be listening to a particular song and just tap to find all files that match, rather than having to manually beat match.

5.2 "Boom-Chick" Rhythmic Retrieval

A slightly more complex way of using rhythm-based information and the proposed interfaces is what we term "boom-chick" retrieval. In this approach rhythmic information is abstracted into a sequence of low frequency (bass-drum like) events and medium-high frequency (snare-drum like) events. Although simple, this representation captures the basic nature of many rhythmic patterns especially in dance music.

The audio signals (drum loops, music pieces) in the database are analyzed into "boom-chick" events using the rhythm-based analysis algorithms described in section 4. The symbolic sensor data is easier to convert into a "boom-chick" representation. Using the EDholak interface is ideal for this application. The first musician taps out a query beat. One piezo sensor represents a "Boom" low event, while a separate piezo sensor represents a "Chick" high event. The query is matched with the patterns in the database and the most similar results are returned. The second musician can then use the digital spoon to browse and select the desired rhythm at a particular tempo. Also time stretching techniques such as resampling and phasevocoding can also be used for changing the tempo of the returned result using the spoon.

5.3 Rhythm-based Browsing

This scenario focuses on browsing a collection of drum samples or musical pieces rather than retrieval. The Radio Drum or the STC-1000 can be used. One axis of the surface is mapped to tempo and the other axis can be mapped to some other attribute. We have experimented with automatically extracted beat strength, genre and dance style for the second axis. The pressure (STC-1000) or the stick height is used for volume control. With this system a DJ can find songs at a particular tempo and style just by placing the finger or stick at the appropriate location. If there are no files or drum loops at the appropriate tempo the system looks for the closest match and then uses time stretching techniques such as overlap-add or phasevocoding to adjust the piece to the desired tempo. Sound is constantly playing providing direct feedback to the user. This method provides a tangible exploratory way of listening to collections of music rather than the tedious playlist-play button model of existing music players.

5.4 Sensor-based Tabla Theka Retrieval

A more advanced scenario is to use the ETabla controller to play a tabla pattern and retrieve a recorded sample, potentially played by a professional musician. This can be used during live performance and for pedagogical applications. There are 8 distrinct tabla basic strokes detected by the ETabla. The database is populated either by symbolic patterns using the ETabla or by audio recordings analyzed similarly to Tindale et al. (2005). This approach can also be utilized to form queries if an ETabla is not available. One of our goals is to be able to automatically determine the type of *theka* being played. *Theka* literally means cycle and we consider 4 types: TinTaal (16 beats/cycle), JhapTaal (10 b/c), Rupak (7 b/c), and Dadra (6 b/c).

5.5 Perforamance Annotation

One of the frequent challenges in developing audio analysis algorithms is ground-truth annotation. In tasks such as beat tracking, annotation is typically performed by listening to the recorded audio signal. An interesting possibility enabled by sensor-enhanced interfaces is to directly provide the annotation while the music is being recorded. This is also important for preserving performance-related information which is typically lost in symbolic and audio representations. Finally, even when listening to music after the fact these interfaces can facilitate annotation. For example it is much easier for a tabla player to annotate a particular *theka* by simply playing along with it on the ETabla rather than having to use the mouse or keyboard.

5.6 Automatic Tabla Accompaniment Generation for the ESitar

The closest scenario to an interactive performance system is based on the ESitar controller. In the playing of the sitar rhythm information is conveyed by the direction and frequency of the stroke. The thumb force sensor on the ESitar controller is used to capture this rhythmic information creating a query. The query is then matched into a database in order to to provide an automatically-generated tabla accompaniment. The accompaniment is generated by matching the rhythmic information into a database containing variations of different thekas. We hope to use this prototype system in the future as a key component of live human-computer music performances.

6 EXPERIMENTS

In order for the rhythm-based matching using dynamic programming the detected "boom-chick" onsest must be accurate. A number of experiments were conducted to determine the accuracy of the "Boom-Chick" detection algorthm described in section 4. The data consists of audio tracks with strong beats which is the focus of this work.

6.1 Data Collection

Three sample data sets were collected and utilized. They consist of techno beats, tabla *thekas* and music clips. The techno beats and tabla *thekas* were recorded using DigiDesign Digi 002 ProTools at a sampling rate of 44100 Hz. The techno beats were gathered from Dr. Rex in Propellerheads Reason. Four styles (Dub, House, Rhythm & Blues, Drum & Bass) were recorded (10 each) at a tempo of 120 BPM. The tabla beats were recorded with a pair of AKG C1000s to obtain stereo separation of the different drums. Ten of each of four "thekas" (meaning beats per cycle) were recorded (Tin Taal Theka (16), Jhaap Taal Theka (10), Rupak Theka (7), Dadra Theka (6)). The music clips were downsampled to 22050 Hz and consist of jazz, funk, pop/rock and dance music with strong rhythm. A large collections of similar composition was used for developing the prototype systems used in the scenarios.

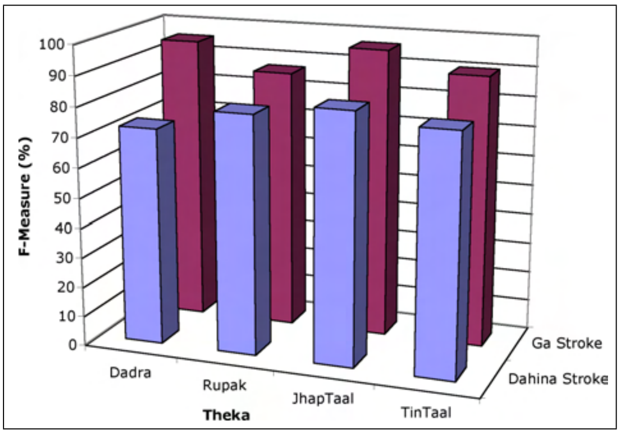

Figure 4: Tabla theka experimental results

6.2 Experimental results

The evaluation of the system was performed by comparative testing between the actual and detected beats by two drummers. After listening to each track, false positive and false negative drum hits were detected seperately for each type ("boom" and "chick"). False positives are the set of instances in which a drum hit was detected but did not actually occur in the original recording. False negatives are the set of instances where a drum hit occurs in the original recording but is not detected automatically by the system. In order to determine consistency in annotation, five random samples from each dataset were analyzed by both drummers. The results were found to be consistent.

The results are summarized using the standard precision and recall measures. Precision measures the effectiveness of the algorithm by dividing the number of correctly detected hits (true positives) by the total number of detected hits (true positives + false positives). Recall represents the accuracy of the algorithm by dividing the number of correctly detected hits (true positives) by the total number of actual hits in the original recording (false negatives + true positive). Recall can be improved by lowering precision and vice versa. A common way to combine these two measures is the so called F-measure defined as (P is precision, R is recall and higher values of the F-measure indicate better retrieval performance):

$$F = \frac{2 * P * R}{P + R} \tag{1}$$

In our first experiment, the accuracy of our algorithm on the Reason drum loops was tested. As seen in figure 5 House beats have almost 99% F-measure accuracy. This is explained by the fact that house beats generally have a simple bass pattern of one hit on each downbeat. For bass drum detection the hardest style was Rhythm & Blues. This can be explained by the largest number of bass hits, which were often located close to each other. The snare drum detection worked well indepenedently of style. One problem we noticed was that some bass hits would be detected as snare hits as well.

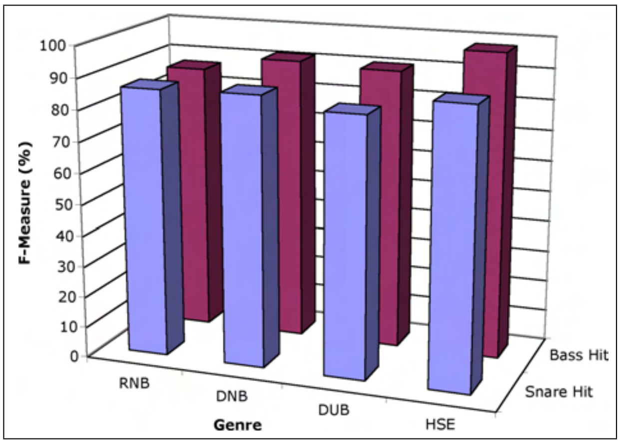

Figure 5: Beat loop experimental results

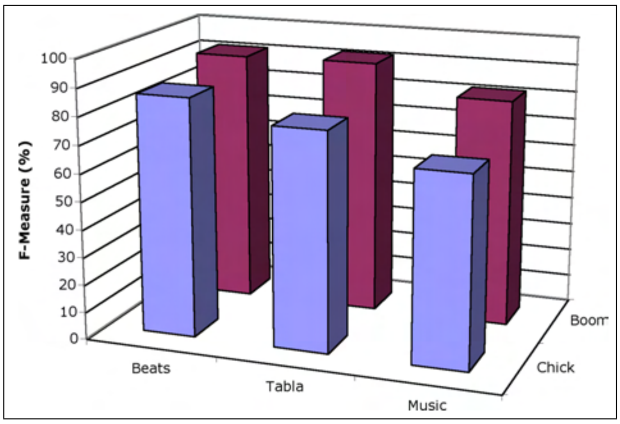

Figure 6: Overall results

The second experiment was conducted on the Tabla recordings. This time, instead of detecting bass and snare hits, ”ga” stroke (the lowest frequency hit on the bayan drum) and ”na” and ”ta” strokes (high frequency hits on the dahina drum) (Kapur et al., 2004b) are detected. From figure 4 it can be seen that the “ga” stroke was detected with high accuracy compared to the dahina strokes.

The final experiment consisted of the analysis of 15 music tracks separated into 3 subgenres; Dance, Funk, and Other. The Dance music results were fairly inconsistent, with a range from 40 to 100% for recall and precision. The bass drum hits were overall more accurate than those found for the snare hits. The Funk music was more consistent, though held the same overall accuracy as the Dance music. The final category, Other, which consisted of Rock, Pop and Jazz tracks was more dependent on the individaul track than the genre. If a pronounced bass and snare drum was present, the algorithm was quite successful in detection. The accuracy of these results was significantly less than those found in Funk and Jazz. As seen in figure 6 the algorithm did not work as well on the these music files as it did on the beats and tabla datasets. This is due to the interference of voices, guitars, saxophones, etc.

Table 1: “Chick” hit detection results

Category	Recall	Precision	F-measure
Rnb	0.844	0.878	0.861
Dnb	0.843	0.891	0.866
Dub	0.865	0.799	0.831
Hse	0.975	0.811	0.886
Average	0.882	0.845	0.861
Dadra	0.567	1.000	0.723
Rupak	0.662	1.000	0.797
Jhaptaal	0.713	1.000	0.833
Tintaal	0.671	0.981	0.727
Average	0.653	0.995	0.787
Various	0.699	0.554	0.618
Dance	0.833	0.650	0.730
Funk	0.804	0.621	0.701
Average	0.779	0.609	0.683

Table 2: “Boom” hit detection results

Category	Recall	Precision	F-measure
Rnb	0.791	0.956	0.866
Dnb	0.910	0.914	0.912
Dub	0.846	0.964	0.901
Hse	0.967	0.994	0.980
Average	0.879	0.957	0.915
Dadra	0.933	0.972	0.952
Rupak	1.000	0.763	0.865
Jhaptaal	0.947	0.981	0.963
Tintaal	0.843	0.965	0.900
Average	0.931	0.920	0.920
Various	0.745	0.803	0.773
Dance	0.823	0.864	0.743
Funk	0.863	0.820	0.841
Average	0.810	0.829	0.819

7 SYSTEM INTEGRATION

In order to integrate the interfaces with the music retrieval algorithms and tasks we developed *IntelliTrance* an application written using the MARSYAS [2] which is a software framework for prototyping audio analysis and synthesis applications. The graphical user interface is written using the *Qt* toolkit [3]. *IntelliTrance* is based on an interface metaphor of a DJ console as shown in Figure 7. It introduces a new level of DJ control and functionality for the digital musician. Based on the standard 2-turntable design, this software driven system operates on the principles of music information retrieval. The user has the ability to analyze a library of audio files and retrieve any sound via an array of preset MIDI interfaces including the ones described in this paper. The functionality of *IntelliTrance* can be found in the main console window. The console offers 2 decks, consisting of 4 independently controlled channels with load function, volume, mute and solo. Each deck has a master fader for volume control, cross fader to control the their amplitude relationship, and

[2] `http://marsyas.sourceforge.net`
[3] `http://www.trolltech.com/products/qt`

Figure 7: *IntelliTrance* Graphical User Interface

a main fader for the audio output. The MIR portion of each track allows for the retrieval of audio files for preview. Once the desired audio file is found from the retrieval, it can be sent to the track for looped playback. *IntelliTrance* offers save functionality to store the audio tracks and levels of the current session and load the same settings at a later date.

8 CONCLUSIONS AND FUTURE WORK

The experimental results show an overall high accuracy for the analysis of audio samples for selected tracks using the "Boom-Chick" algorithm. *IntelliTrance* has a strong focus on music with a pronounced beat therefore these experimental results demonstrate the potential of our approach. The proposed interfaces enable new ways of interaction with music retrieval systems that leverage existing music experience. It is our hope that such MIR interfaces will be commonplace in the future.

There are various directions for future work. Integrating more interfaces into our system such as BeatBoxing (Kapur et al., 2004a) is an immediate goal. In addition, we are building a custom controller for user interaction with the *IntelliTrance* GUI. Another interesting possibility is the integration of a library of digital audio effects and synthesis tools into the system, to allow more expressive control for musicians. We are also working on expanding our system to allow it to be used in media art installations where sensor-based environmental data can inform retrieval of audio and video.

ACKNOWLEDGEMENTS

We would like to thank the student of the CSC 484 Music Information Retrieval course at University of Victoria for their ideas, support and criticism during the development of this project. A special thanks to Stuart Bray for his help with QT GUI design. Thanks to Peter Driessen and Andrew Schloss for their support.

REFERENCES

J. J. Aucouturier, F. Pachet, and P. Hanappe. From sound sampling to song sampling. In *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004.

J. Chen and A. Chen. Query by rhythm: An approach for song retrieval in music databases. In *Proc. Int. Workshop on Research Issues in Data Engineering*, 1998.

G. W. Fitzmaurice, H. Ishii, and W. Buxton. Bricks: Laying the foundations for graspable user interfacesb. In *Proc. Human Factors in Computer Systems*, 1995.

O. Gillet and G. Richard. Automatic labelling of tabla symbols. In *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, Baltimore, USA, 2003.

H. Ishii. Bottles: A transparent interface as a tribute to mark weiser. *IEICE Transactions on Information and Systems*, E87-D(6), 2004.

A. Kapur, M. Benning, and G. Tzanetakis. Query-by-beatboxing: Music information retrieval for the dj. In *Proc. Inter. Conf. on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004a.

A. Kapur, P. Davidson, P. Cook, P. Driessen, and A. Schloss. Digitizing north indian performance. In *Proc. Inter. Computer Music Conf. (ICMC)*, Miami, Florida, 2004b.

T. Machover. Hyperinstruments: A progress report. Technical report, MIT, 1992.

M. Mathews and W. A. Schloss. The radio drum as a synthesizer controller. In *Proc. Inter. Computer Music Conf. (ICMC)*, 1989.

T. Nakano, J. Ogata, M. Goto, and Y. Hiraga. A drum pattern retrieval method by voice percussion. In *Proc. Inter. Conf. on Music Information Retrieval*, Barcelona, Spain, 2004.

H. Newton-Dunn, H. Nakono, and J. Gibson. Block jam: A tangible interface for interactive music. In *Proc. New Interfaces for Musical Expression(NIME)*, 2003.

J. Patten, B. Recht, and H. Ishii. Audiopad: A tag-based interface for musical performance. In *Proc. New Interfaces for Musical Expression (NIME)*, 2002.

J. Paulus and A. Klapuri. Measuring the similarity of rhythmic patterns. In *Proc. Inter. Conf. on Music Information Retrieval (ISMIR)*, Paris, France, 2002.

A. R. Tindale, A. Kapur, W. A. Schloss, and G. Tzanetakis. Indirect acquisition of percussion gestures using timbre recognition. In *Proc. Conf. on Interdisciplinary Musicology (CIM)*, Montreal, Canada, 2005.

G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Trans. on Speech and Audio Processing*, 10(5), July 2002.

K. Yoshii, M. Goto, and H. Okuno. Automatic drum sound description for real world music using template adaption and matching methods. In *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, 2004.

D. Young and I. Fujinaga. Aobachi: A new interface for japanese drumming. In *Proc. New Interfaces for Musical Expression (NIME)*, Hamamatsu, Japan, 2004.