From Real to Virtual: A Comparison of Input Devices for Percussion Tasks

Mike Collicutt **Carmine Casciato** **Marcelo M. Wanderley**
Input Devices and Music Interaction Laboratory (IDMIL)
Centre for Interdisciplinary Research on Music Media and Technology (CIRMMT)
McGill University
Montreal, QC, Canada
mike.collicutt@mail.mcgill.ca casciato@gmail.com marcelo.wanderley@mcgill.ca

Abstract

This paper presents an evaluation and comparison of four input devices for percussion tasks: a standard tom drum, Roland V-Drum, and two established examples of gestural controllers: the Buchla Lightning II, and the Radio Baton. The primary goal of this study was to determine how players' actions changed when moving from an acoustic instrument like the tom drum, to a gestural controller like the Buchla Lightning, which bears little resemblance to an acoustic percussion instrument. Motion capture data was analyzed by comparing a subject's hand height variability and timing accuracy across the four instruments as they performed simple musical tasks. Results suggest that certain gestures such as hand height amplitude can be adapted to these gestural controllers with little change and that in general subjects' timing variability is significantly affected when playing on the Lightning and Radio Baton when compared to the more familiar tom drum and V-Drum. Possible explanations and other observations are also presented.

Keywords: Evaluation of Input Devices, Motion Capture, Buchla Lightning II, Radio Baton.

1. Introduction

The development of new input devices for human movement tracking has paved the way for many new and novel gestural controllers for the production of music. While these devices continue to be created and refined, methods for formally evaluating and comparing different gestural controllers have only recently received significant attention [1] [2] [3] [4] [5]. But since the mapping of gestures to sound as well as the synthesis algorithms used with the controllers can be radically changed, the methods used to evaluate gestural controllers often depend on the context in which they are used; it is therefore interesting to have a specific musical context in mind [1].

In this paper, we decided to focus on percussion tasks: four instruments/devices were selected for a case study in evaluation and comparison: an acoustic tom drum, a Roland V-Drum electronic drum, the Miramax Radio Baton (RB) [7] [8] [9] and the Buchla Lightning II (LII) [9] [10] (see Figure 1). Due to both the LII's and RB's use of sticks and their ability to capture ballistic movements, these controllers may be and actually have been used as percussion instruments in various situations over the last two decades and are among a relatively small group of such devices which have stood the test of time and are still in production today[1].

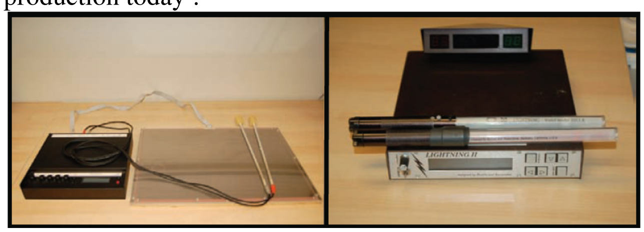

Figure 1.The Radio Baton (left) and the Lightning II (right).

2. Motivation

We present here an analysis of subjects performing simple musical tasks on these four instruments/controllers (herein referred to as 'instruments' for simplicity). These instruments provide a continuum from acoustic instrument (tom drum) to an instrument-like controller (V-Drum), to the RB and LII which can be considered borderline between instrument-inspired and alternate controllers [9] [11]. Each subject's timing accuracy and timing variability were evaluated when subjects performed a series of single strokes on these instruments. In addition, the amplitudes of the subjects' motions were compared across the four devices for the task. It was hypothesized that these

[1] Another recent device that has been increasingly used as a percussion controller is the Nintendo Wiimote. For example, [3] [4] performed an evaluation of the Wiimote along with the Roland Handsonic when performing percussive and other expressive musical tasks. We preferred not to include the Wiimote in this experiment since part of this work is to look on the long-term experience of expert performers with gestural controllers [6].

measurements can provide insight into the following questions:

- Which device performed the best in terms of timing accuracy, and were any differences in timing accuracies statistically significant?
- How are subjects' gestures modified when playing one device compared with another?
- What are user opinions on the efficacy of the LII and RB as percussion instruments?

3. Experiment

Three subjects from the classical percussion program in the Schulich School of Music, McGill University were selected for the experiment. Subjects were asked to perform two basic percussion tasks: single strokes (SSF), and double stroke rolls (DSR) in this order, both at forte volume and to a 120 bpm click track. Each subject performed between 16 and 32 strokes until they were satisfied with their performance. Subjects were given up to 10 minutes during setup of the different instruments to warm up and become familiar with them. This warm up time was considered sufficient to become familiar to the current instrument, while forgetting specific techniques used in any previous one. The exercises were completed on the tom drum, V-Drum, RB, and LII (in this order) for all subjects. The three gestural controllers in this study were mapped to a single tom drum sample from the V-Drum sound engine. The strike velocity in the RB and LII were mapped to the volume of the sample. The volume of the tom drum and sound sample was subjectively adjusted by listening to the real-time sound during setup.

A Polhemus Liberty 8 movement tracker was used to capture the subjects' actions during the tests. The sensors were placed on the back of each hand, on the forearms, biceps, back, and top of the head. Subjects were also recorded during the tests via a JVC GR-HD1 camera with a shutter speed of 1/1000 sec at 29.97 fps.

4. Analysis

Analysis of the subjects' performances was completed using the Polhemus data alongside the video recordings which were parsed, aligned side by side, and synchronized to allow for convenient comparison of techniques that subjects used (Figure 2).

The Polhemus system tracks each sensor with 6 DoF (3 spatial, 3 rotational). Since the majority of subjects' motions were observed to be in the vertical direction of the forearms and hands for these exercises, only this data was used in the analysis of timing and amplitudes of gestures.

4.1 Assumptions & Sources of Error

Ideally, the Polhemus sensors would be attached directly to the striking element of the device. However, because they are wired, attaching them to the striking element would prove too intrusive to the subjects' playing. Thus the minima in the hand heights were assumed to correspond to strikes on the interface.

Figure 2. A screen-shot from the video recording of Subject 1 during one of the tests. From left to right are the Subject playing: the tom drum, V-Drum, RB, and LII.

In the majority of Subject 1's task performances however, it was clear that the minima of the hand heights did not actually correspond to a strike, but rather a strike occurred slightly before a minima at what looks like a slight 'glitch' in the hand height data. This is most likely a result of the stick rebounding off of the playing surface. In the case of the LII, this effect was also observed to exist, and was most likely from the subjects' sudden halting of their arm movements, resulting in a slight jitter. This effect is illustrated in Figure 3, however it was not observed in the other subjects (Figure 4): for Subjects 2 and 3, the minima do correspond to strikes.

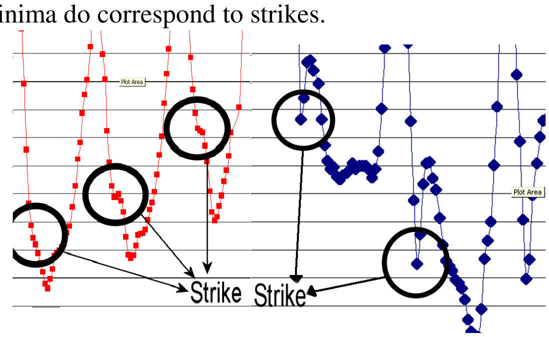

Figure 3. Polhemus hand height positions of Subject 1 during strikes on the LII (left) and tom drum (right). Here the strikes do not correspond to minima, but rather to the circled points.

When searching for minima in data, it is clear that the actual minima might light somewhere between two data points, and so there is an inherent uncertainty in the location of each strike of ±16.5ms, as the sampling period for the Polhemus data is 33ms for all markers[2]. This uncertainty in time also transfers to an uncertainty in hand height, in particular, when picking out strikes and maxima of the data. An estimate for the uncertainty in hand height of 0.1cm has been assumed. This value is believed to still

[2] This relatively low sampling rate had to be used due to a limitation in the protocol used to send the Polhemus data through the network.

be conservative, but closer to the true uncertainty, which could not be determined without further tests.

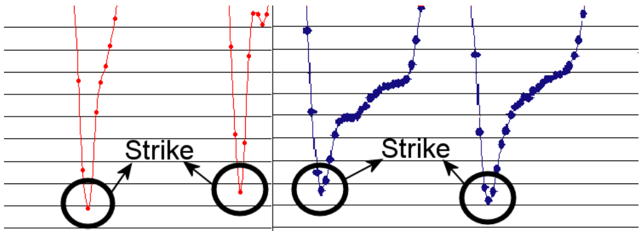

Figure 4. Polhemus hand height positions of Subject 3 during strikes on the LII (left) and tom drum (right). Subject 3 does not exhibit the 'jitter' like Subject 1.

It was also assumed that the first strike by a subject was timed perfectly to the click track. This was necessary because the click track was not synchronized with the Polhemus data or the video recordings, therefore creating an unknown constant offset in the timing data. This offset is presumed to be small, but variable between each exercise and each Subject.

4.2 Data Analysis

Motion capture data was obtained for the subjects performing the SSF and DSR musical tasks. Subjects' motions in this work were analyzed while performing the single strokes task only. A total of 16 consecutive strokes were analyzed for each subject on each of the instruments.

4.2.1 Timing Accuracy and Variability of Strokes

The acquisition of minima from the hand height data was essentially done by a careful analysis of the hand height plots along with video recordings that made it fairly easy to pick out strike locations in the data by hand[3].

While this method does not account for the potential latency of each device, the latency is assumed to be constant for an instrument[4]. Since performers were given a substantial warm up period beforehand, it is assumed that the Subjects took the inherent latency into account when performing the tasks and attempted to match the instrument's sound output to the click track. Thus the assumption on the first strike means that the following analysis is done relative to a click track that is simply offset from the one in the experiment by an amount equal to the device's latency. Therefore as long as the latency is constant for each instrument, it is fully taken into account in the data.

4.2.2 Amplitude Variability of Stroke Gestures

The acquisition of data to determine the amplitude of gestures was done in the same way as for the acquisition of timing data. A simple formula relating the hand height position of a given strike, the maximum amplitude of the hand attained between this strike and the next, and the height of the next strike was devised and is given by:

$$Amp[n] = (z_{\max}[n] - z_{\min}[n-1]) + (z_{\max}[n] - z_{\min}[n]) \quad (1)$$
$$Amp[n] = 2z_{\max}[n] - z_{\min}[n-1] - z_{\min}[n]$$

Where $z_{max}[n]$ is the peak hand height between two strikes for a given hand, and $z_{min}[n-1]$and $z_{min}[n]$ are the hand heights at the location of the two strikes surrounding the maximum for a single hand. This gives a meaningful representation of the amplitude of the performers' gestures while still maintaining simplicity and ignoring fine-grain details dependent on the subjects' playing styles.

Equation 1 only shows changes in the gestures of performers in a relative sense; it does not consider how the hand heights might change over time relative to some fixed marker like the ground. A qualitative analysis of how the absolute hand heights change is discussed in Section 5.2.

5. Results

Calculations of the mean timing accuracy of strokes and amplitude variation of gestures were completed for each of the subjects playing a SSF exercise on each of the four instruments.

5.1 Timing Accuracy of Strikes

The accuracy of the strikes was first characterized in terms of the mean timing error relative to a 120 bpm click with a variability characterized by the standard deviation of these timing errors. Uncertainties in the means were substantially lower than the standard deviations, and so are not shown. Results are shown in Figure 5.

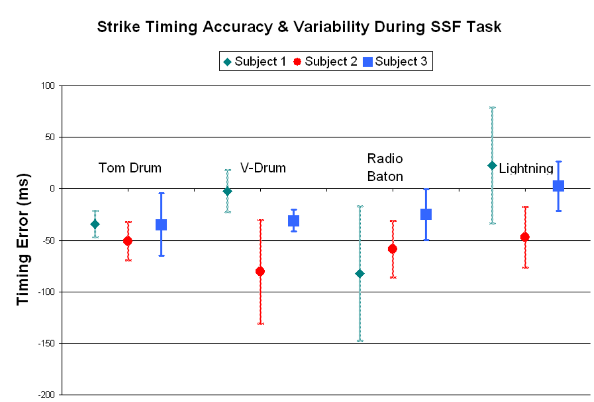

Figure 5. Mean timing accuracy for each subject playing single forte strokes on each of the four instruments. Error bars are the standard deviations of timing errors. A negative timing error indicates an anticipation of the click track.

An analysis of the timing data was completed using one-way ANOVA [12]. Results showed that the differences in timing accuracy were very significant ($p<0.001$) for Subjects 1 and 3 and significant ($p<0.05$) for Subject 2. A Tukey's Honestly Significant Difference (HSD) test was applied to the ANOVA data for each subject to see where

[3] Future work with this data will automate this process.
[4] Latency estimates by long-time performers are given in [9].

the significance lies. Table 1 shows the results of the Tukey HSD test for each subject.

It was expected that there would be no significant difference between timings of the tom drum and V-Drum, as the V-Drum is designed to mimic a real tom drum. It was also expected that there would be a significant difference in timing errors between the LII and any of the other instruments, due to the lack of playing surface with the LII. Table 1 however shows that only Subject 3 experienced a significant difference between the LII and each of the other instruments. Subject 1 only showed a difference between the LII and RB, and Subject 2 experienced a difference between the LII and V-Drum.

Table 1. Results of Tukey HSD Test. "SIG" represents a significant, non-statistical difference between the mean timing errors for a subject. "NOT SIG" represents differences in the mean timing errors that can be attributed to statistical variability.

Comparison	Subject 1	Subject 2	Subject 3
Tom & V-Drum	NOT SIG	NOT SIG	NOT SIG
V-Drum & RB	SIG	NOT SIG	NOT SIG
RB & LII	SIG	NOT SIG	SIG
V-Drum & LII	NOT SIG	SIG	SIG
Tom & LII	NOT SIG	NOT SIG	SIG
Tom & RB	SIG	NOT SIG	NOT SIG

It is also interesting to note that the rows for 'Tom & RB' and 'V-Drum & RB' present similar results. This is also true for the 'Tom & LII' and 'V-Drum & LII' rows with the exception of Subject 2. This was expected as the nature of the V-Drum is to closely mimic a traditional tom, and so results should be the same whether one is going from either a tom drum or V-Drum to something much different, like the RB or LII. Note, on the other hand, that the timings for the various subjects differ. This means that subjects were consistent when comparing devices (e.g. 'V-Drum & LII' and 'Tom & LII'), but their timings were not necessarily similar.

5.2 Amplitude Variation and Regularity

The peaks in the hand height data were first extracted from the Polhemus tracker for each hand and used in conjunction with the strike data used previously. Equation 1 was then applied for both hands, and the results were characterized using the mean and standard deviation for each hand on each instrument. The mean amplitudes for each hand were combined to give a single value for each instrument. Results are shown in Figure 6.

With the exception of Subject 2, the subjects did not appear to modify their gestures very much when playing the different devices. Typically one would expect very similar gestures to be made between the tom drum and V-Drum, since both employ drum sticks for striking, and have roughly the same rebound properties. It was observed from the video data that Subject 2 used much larger gestures on the V-Drum than the tom drum. In addition, Subject 2 used a different grip (French grip) than the other Subjects, who both used German grips for all of the instruments. These factors could explain the observed differences in timing variability and gesture amplitude measurements for Subject 2 on the tom and V-Drum. A significance test using ANOVA showed significant differences in gesture amplitude for Subjects 2 and 3 ($p<0.001$). Subject 1 showed insignificant differences ($p<0.6$). Once again, Tukey's HSD test was used to find where the significant differences lie for Subjects 2 and 3. Results are presented in Table 2.

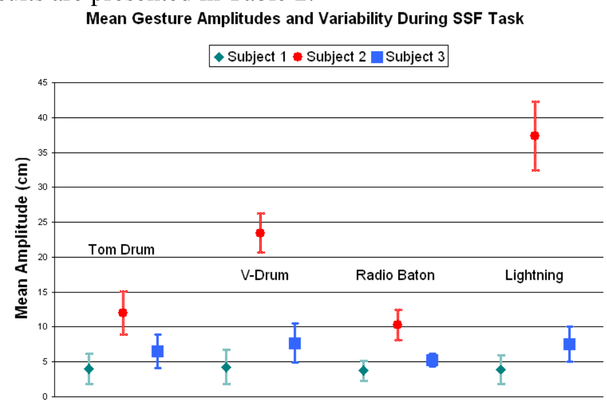

Figure 6. Mean gesture amplitudes for each subject playing single forte strokes on each of the instruments. Errors bars are the standard deviations of the gesture amplitudes. Data is also averaged between left and right hands.

A few consistent observations arise out of the analysis. Again, with the exception of Subject 2, the player's hand heights were not significantly affected when transferring from the tom to the V-Drum, which was expected. Also expected was that the 'Tom and LII' entry is the same as the 'V-Drum & LII' entry for all subjects for the same reasons as mentioned in Section 5.1. This however was not the case with the 'V-Drum & RB' and 'Tom & RB' entries for Subject 3.

Table 2. Results of Tukey HSD Test. "SIG" represents a significant, non-statistical difference between the gesture amplitudes for a subject. "NOT SIG" represents differences in the gesture amplitudes that can be attributed to statistical variability.

Comparison	Subject 1	Subject 2	Subject 3
Tom & V-Drum	NOT SIG	SIG	NOT SIG
V-Drum & RB	NOT SIG	SIG	SIG
RB & LII	NOT SIG	SIG	SIG
V-Drum & LII	NOT SIG	SIG	NOT SIG
Tom & LII	NOT SIG	SIG	NOT SIG
Tom & RB	NOT SIG	SIG	NOT SIG

The Polhemus data also showed that the envelope of subjects' movements stayed mostly flat while performing a task on one of the devices. However, it was observed that for Subjects 1 and 3, their non-dominant hand was consistently up higher than their other hand while playing the tom drum and V-Drum. This effect disappeared however when those subjects played on the RB and LII.

5.3 Double-Stroke Rolls Exercise

It is interesting to note that the same difference in hand height was observed for Subject 3 when performing DSRs. Subject 1 however showed this difference on the tom drum and the RB, but not on the V-Drum.

Performance of this exercise was poor when subjects used the RB or LII. The audio from these exercises show many missed strike triggers on both instruments. This is discussed further in the following Section.

5.4 Post-Experiment Subject Comments

Following the completion of the SSF and DSR tasks on the four instruments, subjects were asked to rate on the V-Drum, RB, and LII in terms of a number a features relative to the tom drum (assumed rating of 5 for all categories). A summary of subjects' responses are shown in Table 3.

Table 3. Summary of post-experiment subject comments (1 is poor and 5 is excellent).

Subject	1	2	3	Average
Timing Accuracy				
V-Drum	4	5	5	4.67
Radio Baton	3	4	3	3.33
Lightning	2	3	1	2
Timbre Controllability				
V-Drum	4	4	5	4.33
Radio Baton	2	3	1	2
Lightning	3	3	2	2.67
Responsiveness				
V-Drum	4	4	5	4.33
Radio Baton	2	3	3	2.67
Lightning	2	2	1	1.67
Expressivity				
V-Drum	4	3	5	4
Radio Baton	2	3	2	2.33
Lightning	2	3	2	2.33
Acoustic Likeness				
V-Drum	4	5	4	4.33
Radio Baton	3	3	2	2.67
Lightning	2	1	2	1.67
Playability				
V-Drum	5	5	5	5
Radio Baton	3	3	3	3
Lightning	2	1	1	1.33

Overall, the V-Drum was considered by far the best facsimile of an acoustic drum, followed by the RB and then the LII. It is interesting to note that subjects ranked the LII higher than the RB for Timbre Control, and equally good for Expressivity. Furthermore, the impressions about timing accuracy are especially interesting as they contradict the results of the single stroke condition.

6. Discussion

Results showed no significant differences in timing accuracy between the Tom Drum and V-Drum for all subjects. Data from Table 1 and Figure 5 suggest that in general timing is significantly better in the LII. This seems counter-intuitive, as the LII has no physical interface to strike. Furthermore, subjects rated the LII worst in terms of timing expressivity (see Table 3), probably because of the LII's poor performance during the DSR exercise.

One must be careful before concluding that the results of the timing experiment done in this work can be generalized. The assumption that the subjects' first strike was timed perfectly introduces an unknown offset in the timing data that is different for each instrument and subject. While this offset is assumed to be relatively small, it nonetheless could confound the timing results obtained.

The variability does not suffer from this assumption however. Analysis of the subjects' timing variability shows that the tom drum has the smallest variability on average, which is to be expected. The LII had the next lowest average variability across the 3 subjects, closely followed by the V-Drum and lastly the RB. The low variability of the LII data is surprising, as it is the most dissimilar from an acoustic tom drum. One possible explanation is that because of the lack of a playing surface, performers are forced to halt their movements, rather than 'letting the stick do all the work' [1]. Hence gestures used with the Lightning appear to be fundamentally different than those used with the other instruments. Further research is needed to develop an explanation for this result.

From Table 2 it appears that in general subjects (with the exception of Subject 2) do not have to adjust their gestures significantly when moving from acoustic instruments to percussion-based controllers when playing single strokes. Results show that the majority of comparisons between the "traditional" instruments (tom and V-Drum) showed no significant adaptation of gesture amplitude when compared to performances on the RB or LII. However, when one looks at hand heights in an absolute sense, differences are found when going from the tom drum or V-Drum to the two gestural controllers in either exercise. This suggests that subjects approached the gestural controllers differently in the SSF and DSR tasks.

It is clear from the videos captured during the performances that subjects indeed played the RB and LII differently during the double stroke rolls exercise by exerting more effort and energy to create the 'roll' effect due to the lack of significant rebound on the foam-tipped RB sticks and lack of any playing surface at all on the LII. Despite this extra effort, audio results from the tests show that subjects could not perform the roll with the RB and LII nearly as well as the tom and V-Drum as the trigger algorithms in the RB and LII could not handle the stroke density of the exercise. This is reflected in the post-experiment Subject comments (see Table 3). Subjects rated the RB and LII much lower than the V-Drum in terms of timing expressivity, responsiveness, and playability. It is also interesting to note that the RB and LII scored substantially worse than the V-Drum in terms of timbre control and expressivity in light of the fact that all gestural controllers were only controlling the volume of the same tom drum sample. It therefore appears that the subjects have a general negative view of the RB and LII which influenced their ratings of timbre control and expressivity for these instruments.

As the work in [3] [4] and [9] show, one can obtain a more complete picture of the performance of musical controllers when the instruments are studied through a number of different approaches. A multi-faceted approach is important in the evaluation of musical devices as musical instruments are often be quite complex and subtle; offering fine-grain control over a number of different parameters. For example, in [3] and [4] subjects using the Nintendo Wiimote and Roland Handsonic in percussion-based tasks showed no significant differences in timing error between the two controllers. However, post-experiment interview data showed that subjects greatly preferred the Handsonic to the Wiimote, mainly due to the lack of physical feedback with the Wiimote and the difficulty level of performing faster rhythms. On the other hand, however, subjects did enjoy the intuitive nature of the Wiimote as well as its portability.

Similarly in this work, an analysis of the timing accuracy, variability, gesture amplitude, and interview data has shown contrasting results in particular with respect to the timing accuracy and variability of the LII. While it appears that percussionists can easily play the simple single stroke exercise, the double stroke rolls and subject comments suggest that there is a need for extra practice with the RB and LII so that percussionists can perform more developed exercises on them.

7. Conclusions and Future Work

This work has performed both a quantitative and qualitative analysis of subjects' motions while playing simple percussion tasks on four percussion "instruments", from real to virtual: a standard tom drum, Roland V-Drum, Radio Baton, and Buchla Lightning II. In addition, post-experiment comments from subjects have helped to provide a clearer picture of the overall efficacy of the instruments. Analyses of timing accuracy and variability as well as gesture amplitude have been completed for the SSF exercise. Analysis of the timing variability of strokes show that next to the tom drum, the LII was the least variable of the 3 other instruments. The reasons for the better timing variability of the Lightning could lie in the different ways that users control their movements when playing an instrument with no surface to strike. It was also shown that for simple exercises, gesture amplitudes were not significantly different when playing each of the instruments. However, there did appear to be some modification in the relative left and right hand heights when subjects went from the traditional instruments to the electronic ones.

To support the hypothesis developed to explain why timing accuracy and variability were best in the LII, further experiments and research on the gestures used when playing the LII are necessary. It would also be interesting to perform this study with subjects executing a number of trials to avoid any performer-dependent inconsistencies that may arise.

Lastly, it would also be interesting to investigate different percussion exercises and see how well they can be adapted to these new controllers.

8. Acknowledgements

We would like to thank Bruno Giordano for suggestions and help with the data analysis in early versions of this manuscript as well as Fernando Rocha and Fabrice Marandola for percussion exercise suggestions and comments and all of the subjects who participated. Thanks also to the anonymous reviewers for their insightful comments. This work is supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada (Discovery Grant) to the third author.

References

[1] M. Wanderley and N. Orio, "Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI," in *Computer Music Journal*, vol. 26, no. 3, pp. 62-76, 2002.

[2] D. Isaacs, *Evaluating Input Devices for Musical Expression*, B. Inf. Tech. Thesis, University of Queensland, 2003.

[3] C. Kiefer, N. Collins, and G. Fitzpatrick. "HCI Methodology for Evaluating Musical Controllers: A Case Study," in *NIME Proceedings, 2008.*

[4] C. Kiefer, N. Collins, and G. Fitzpatrick. "Evaluating the Wiimote as a Musical Controller," in *Proceedings of the ICMC,* 2008.

[5] D. Stowell, M. Plumbley, and N. Bryan-Kinns. "Discourse Analysis Evaluation Method for Expressive Musical Interfaces," in *NIME Proceedings, 2008.*

[6] C. Casciato and M.Wanderley. "Lessons from Experienced Gestural Controller Users" In *Proceedings of the 4th International Conference on ENACTIVE Interfaces*, 2007.

[7] J. Paradiso, "American Innovations in Electronic Musical Interfaces: Baton Interfaces," *New Music Box,* 1999.

[8] R. Boulanger and M. Mathews, "The 1997 Mathews Radio-Baton and Improvisation Modes," in *Proceedings of the ICMC,* pp. 395-398, 1997.

[9] C. Casciato, *On the Choice onf Gestural Controllers for Musical Applications: An Evaluation of the Lightning II and the Radio Baton*, Master's Thesis, McGill University, 2007. Available online: http://idmil.org/publications.

[10] R. Rich, "Buchla Lightning MIDI Controller," in *Electronic Musician,* vol. 7, no. 10, pp. 102-108, 1991.

[11] E. Miranda and M. Wanderley, *New Digital Musical Instruments: Control and Interaction beyond the Keyboard*. Wisconsin: A-R Editions, 2006, pp. 19-30.

[12] T. Hill and P. Lewicki, *Statistics Methods and Applications*. Tulsa, Oklahoma: StatSoft, 2007, pp. 41-56.

[13] J. Delecluse, *Methode de Caisse-Claire*. Paris: Alphonse LeDuc, 1969, pp. 23-30