Towards an evaluation methodology for digital music instruments considering performer's view: a case study

Jerônimo Barbosa[1], Filipe Calegario[1], Francisco Magalhães[1], Giordano Cabral[2], Veronica Teichrieb[1], Geber Ramalho[1]

[1]Centro de Informática – Universidade Federal de Pernambuco (UFPE)
50.740-560 – Recife – PE – Brazil

[2]DEINFO – Universidade Federal Rural de Pernambuco (UFRPE)
52.171-900 – Recife – PE – Brazil

`{jbcj,fcac,fpms,vt,glr}@cin.ufpe.br, giordanorec@gmail.com`

Abstract. *With the increasing number of initiatives concerning the design of new Digital Musical Instruments in the recent years and with the perceivable lack of methods for evaluating them, this work proposes a process for evaluating those devices considering the performer's perspective. It also presents a case study which describes and evaluates its practical application. The proposed process could be also utilized as a very useful step when designing new DMIs, providing a structured tool for incremental development of prototypes based on user feedback.*

1 Introduction

The desire of using computers in the process of making novel musical instruments is old and may be perceived in numerous attempts through the history of Computer Science, like MUSIC, by Max Mathews in 1957, The Hands, by Michel Waisvisz in the 1980s, Hyperinstruments, by Tod Machover in 1986, and others [12][13].

In recent years, the number of researches concerning Digital Musical Instruments (DMIs) has increased, as can be seen in several works [3][17][19]. DMIs can be defined as musical instruments that allow the separation of their control interface and sound synthesis as well as their relation by using some mapping strategies [11]. In contrast with acoustical musical instruments, which are limited by their physical ways to produce sound, DMIs designers can think the physical design with more freedom due to the fact that the sound is produced digitally by an independent module of the system [12], providing a vast range of musical possibilities for musical composition and performance.

DMIs are composed by the control interface and the sound synthesis parts [11]. The controller is a device (or can be understood as a set of sensors) that transforms a value from nature (analogical input) into a computable number. In the DMI design, this value can be mapped into a variety of musical variables such as frequency and volume, for example.

The increasing importance of DMIs in recent researches has aroused a large number of questions that are still open. One of the most important is: how this kind of system can be objectively evaluated?

This importance becomes clear when we observe the nature of DMIs; part of them is originated from experimentation and continuous development of prototypes. As mentioned by Wood [28], a critical ingredient for designing systems is "understanding potential users", what involves observing them and analyzing the information collected about their work - characteristics that should be considered in good evaluation methods. In this way, creating structured evaluation methods would help this process, providing better analysis tools and expanding possibilities for getting user's feedback.

Despite the importance of this question, researches about this issue are rare [2]. One may glimpse this lack of bibliographical content analyzing specialized literature. For example, taking into account the proceedings of the International Conference on New Interfaces for Music Expression (NIME), it's remarkable the few number of papers concerning the use of any kind of formal evaluation, as shown in Figure 1.

Evaluation type	NIME conference year		
	2006	2007	2008
Not applicable	*8*	*9*	*7*
None	18	14	15
Informal	12	8	6
Formal qualit.	1	2	3
Formal quant.	2	3	3
Total formal	*3 (9%)*	*5 (19%)*	*6 (22%)*

Figure 1: Analyzing NIME publications in the years of 2006, 2007 and 2008 according to the kind of evaluation proposed [2]

Although the analysis being incomplete, it seems quite representative, since NIME is one of the most important events in this domain, and has been contributing for the development of the area for more than 10 years.

1.1 Evaluation of DMIs

One important aspect that must be considered in an evaluation process is the variety of stakeholders involved in the use, conception, perception and even commercialization of a musical device. As mentioned in O'Modhrain's work [3], this lead us to a more generic concept of evaluation, in which the whole evaluation process should consider: (a) Performer's view - How effective is the relationship between performer and device in a manner that the second allows the first to concretize all his musical intentions; (b) Audience's view - How effective is the relationship between performer and device in a manner that could affect sensitively the ones who watch the performance? (c) Manufacture's view - How effective is the system under a commercial perspective? In this way, when designing a new DMI, it is necessary to consider all these perspectives, before creating specific methodologies for evaluating each one of these characteristics.

An initial and natural approach for evaluating DMIs is to use techniques from Human-Computer Interaction (HCI) domain. In this context, Wanderley et al. [16] propose a quantitative task-based approach, in which potential performers were asked to perform simple musical tasks. Unfortunately, the direct reuse of HCI techniques may be problematic in some cases. For instance, "talk-aloud" protocols and others based on models of human cognition are not perfectly suitable due to their possible influence in musical experience. Wanderley's approach itself was criticized by this reason [2].

Considering this problem, Stowell et al. [2] propose a structured qualitative evaluation methodology based on Discourse Analysis. This work is very important since it provides valuable information about the interaction between musicians and new musical interfaces. Besides, it highlights the importance of using qualitative approaches instead of quantitative ones, due to the subjectivity of a musical experience, that "requires alternative ways for evaluating them" and are hard to be achieved through techniques like questionnaires.

Despite the work of W. Hsu and M. Sosnick [22] being based mainly on quantitative evaluation, their evaluation methodology proposal brings an interesting idea of using individual user sessions as rehearsals and the group session as a short performance, giving the user a specific objective for deeper exploration of the musical system being tested. The expectation is that this approach provides a more concise and richer information set about the musical device, due to the target performance.

In other hand, Birnbaum [4] proposes a "holistic and informative" way to visualize these devices through the building of a 7-axis graph that considers important aspects of DMIs, technique called Dimension Space Analysis. It is a very powerful in the context of comparison due to its visual nature and their systematic definition and isolation of relevant DMIs characteristics. However, it is important to remark that the method utilized for plotting these graphs is still an open issue and the work does not provide any practical case study on how the technique could be applied.

Based on previously described attempts, this work proposes a combination of ideas, techniques and methods in order to provide a more complete approach on evaluating DMIs considering the performer's view (henceforth called user), because it was appointed as the most important stakeholder in the process [3].

For that, we have merged the above mentioned previous attempts in a single and unified qualitative process, in which the idea of using solo sessions and group sessions, with later application of discourse analysis, from Stowell et al. [2] was enhanced with the concepts of rehearsal and performance from W. Hsu and M. Sosnick [22].

Finally, Birnbaum [4] dimension space analysis was incorporated to the methodology to obtain an easy way of visualizing the system characteristics and facilitate the comparison with other systems according to a set of parameters. Furthermore, we also propose to use the discourse analysis results for plotting the graph.

2 Methodology

The process proposed is divided in three steps: *data collection*, *data analysis* and *data visualization* - each one using different techniques described as follows.

2.1 Data collection

The first step aims to collect information about the user experience with the device and will serve as base for the following stages.

The data collection step will be done through two distinct moments:
- Solo sessions: Moment of individual rehearsal, divided in two sub stages: *Free-Exploration,* where the user is invited to explore the musical devices without any interference from the researchers; and *Guided Exploration*, where the

researchers give some tips on how the system works. For both cases, *semi-structured interviews* [5] are used for obtaining the data;

- Group session: Moment in which all participating users meet the researchers for: (1) present their individual performances; (2) share their experience in using and watching others performing with the system. For that, a *focus group* [7] was used for collecting the data for later analysis.

Born in the context of sociology and psychology [6] and widely used as method of traditional HCI [5], both semi-structured interviews and the focus group present a good potential for DMIs evaluation due to its qualitative nature, more suitable for a complex and subjective context like musical interaction. Furthermore, they were already used in previous works in literature [2], what makes them a suitable option for collecting data.

Both solo sessions and the first part of the group session (performances presentation) were considered for evaluating the performer's view. However, despite its importance, the focus group analysis will not be described here due to the fact it aims the audience's view and it will be used as basis for future works.

It also should be highlighted that the structure of the interview was based on the dimension space axis criteria, whose technique is used in the data visualization stage and described in the following sections. Besides, the interview should be applied only after the musical experience. As mentioned before, interviewing during the user-device interaction is unsuitable because speech may interfere directly the musical experience. It is recommended that all interactions (both moments) should be recorded. It could be used to stimulate comments during the interviews and for later analysis.

2.2 Data analysis

The second step has the purpose of analyzing the information acquired in the interview. For that, it was used a method known as *Discourse Analysis* (DA), which is an analytic tradition widely used in linguistics and social sciences, that allows us to analyze the discourse (written texts, discourses, conversations, and others) looking at patterns across texts as well as the social and cultural contexts in which the texts occur [8]. It was chosen for the process here proposed because of its successful application in a previous work [2], since (1) it provides a structured method for the evaluation of the interviews, aiming at more analytic reliability; and (2) it takes into account social structures presented on the analyzed material, what is utterly suitable due to social aspects that influence the whole interaction in musical and technological contexts.

DA has five steps:
a. Transcription - Step in which all speech material is transcribed to text. This is necessary because DA always uses written text as raw input data for its process;
b. Free association - Step in which impressions are noted down in a free way, for later analysis;
c. Itemization - Step in which the whole text is broken into small pieces, where the most common elements in discourse are recognized, using as far as possible the terminology utilized by the participants;
d. Reconstructing user's world - Step in which the analyzer builds a representation of each experience by the perspective of the respective participant, based on this list of most commonly used elements and their relationship;

e. Examining context - Finally the last step, in which the representation is compared and analyzed.

2.3 Data visualization

The last stage of our process aims to show the information in a clearer and more intuitive manner, providing a better way to visualize and, consequently, analyze the results. For that, we use a technique called *Dimension Spaces Analysis*.
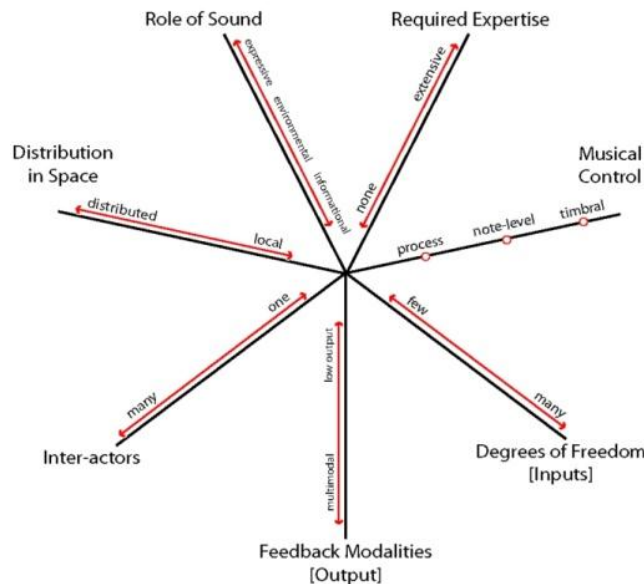
Figure 2: The 7-Axis Dimension Space created by Birnbaum et al. [4]

Dimension Space Analysis, which has already been used in several different contexts [4], is a technique in which the central idea is to build a multi-dimensional graph that allows us to visualize characteristics of the studied element, in which each axis represents a *continuum* of different and independent aspects of the system.

A sketch of its utilization in the context of DMI was proposed originally by Birnbaum et al. [4] (Figure 2) and it could be described in function of the following axis:

a. Required expertise - Describes how much previous practice and training the performer needs for working with the instrument;
b. Musical control - Describes how much control the performer owns about the musical output;
c. Degrees of freedom - Describes how much input control the system provides for the performer;
d. Feedback modalities - Describes how much real-time feedback (e.g. visual, auditory, tactile, kinesthetic) the user receives from the system;
e. Inter-actors - Describes the number of people that could be involved with the musical experience;
f. Distribution in space - Describes the physical area necessary for interacting with the system;
g. Role of sound - As described by Birnbaum et al., this axis uses "Pressing's categories of sound roles in electronic media". This scale ranges from

artistic/expressive (all kinds of music and songs), to informational (speech, alarms and so on), passing through environmental (noise of animals, wind, industrial noise, among others) [13].

These axis were also used as basis for the applied interviews structure, aiming to facilitate the processing of building the dimension space.

It is important to notice that the previous work does not cover how the graph should be plotted in the context of DMIs. For that, it is proposed the use of previous steps (data collection - with interviews and focus group - and data analysis - with discourse analysis).

2.4 Prototype

In order to validate and refine the proposed evaluation methodology we developed a prototype and performed some case studies, as described in next sections. It is important to highlight that the intention of this paper is not to defend the quality of this particular application, but instead to explain the process by means of this example.
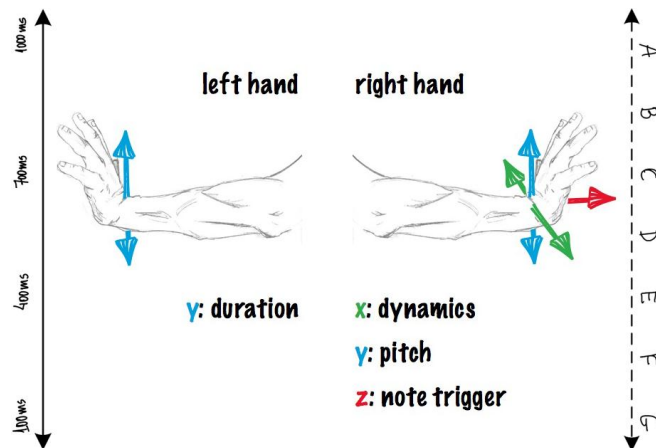
Figure 3: The repertoire of gestures used in the prototype and their mapping into pitch and duration scales

The prototype is composed by a controller (Microsoft Kinect Sensor [23]), a music variables mapper (re.scalla [25]) and a sound synthesizer (SimpleSynth [26]). Based on Wanderley et al.'s classification for DMIs [16], it can be categorized as an alternate controller.

As illustrated in Figure 3, the gestures repertoire covers only the movements of hands and uses their position to influence musical variables concerning a note flow being played.

The movement of the user's hand in the Y-axis will be discretely mapped into a musical scale and the left hand position will influence the note duration, as can be seen in Figure 4. When the right hand Z-value reaches a certain threshold (the further it becomes to the Kinect sensor), a note flow starts to play in a specific beat determined by the left hand Y-position and with the pitch specified by the right hand Y-mapping.

For mapping both hands positions into pitch, duration and note trigger control, the re.scalla [25] library was used. The normalized (from 0.0 to 1.0) parameters related to the hands positions were mapped into musical variables: (1) discrete notes scale and (2) continuous duration scale, outputting the notes flow to SimpleSynth.

The Microsoft Kinect Sensor [23] used in this work is based on an infrared light projector and an infrared camera allowing a depth view of the scene captured by the sensor as well as the tracking of objects in 3D space.

For further information about how users interact with the system (also shown in Figure 4), a demonstrative video is available in internet [27].

Figure 4: A user interaction example

3 Case Study

The experiment was developed with 4 people, aged 18-25, all having some familiarity with technology and music, but none being professional musicians. Each participant had two different moments for rehearsal, the free and the guided one, followed by a formal presentation, when each one showed a short artistic performance using the features they liked most in the system.

This data acquisition stage took in average 40-50 minutes per solo session for each participant, in which 10-20 minutes were semi-structured interviews and 2 hours the performances followed by the focus group, totalizing around 7 hours and half for this stage. The data analysis was applied to the collected material through DA methods, followed by the data visualization stage. Both took about 5 hours per participant, totalizing around 20 hours. The original collected material (in Portuguese) is available for download on the internet [27].

3.1 Reconstructing user's world

Below it is presented a paragraph for each user as a result of DA's *reconstructing user's world* step, previously described in Methodology section.

3.1.1 User 1

User 1 was the user who had the lowest musical knowledge among the interviewed users. He stated that the way of making music was "easy and fast to catch", even when he did not know exactly how the system works (free-exploration stage), but the "noise" produced was "boring". Besides he could not say which musical variable he was controlling, he complained about the lack of visual feedback of the system, "I wish I could see me (...) as a little help for guiding me to know what I am doing". He also believes that it is possible to become a virtuoso by training. He felt physically tired by the end of the first moment. This has not happened at the second time due to, according to him, the "necessity of building something to show" for the performance.

3.1.2 User 2

User 2 is familiar technology, constantly listens to music and despite playing acoustic guitar for years, does not consider himself as a professional musician. He described his relation with the system making an analogy with a tamed animal: "the system tamed me,

when I was supposed to tame it". This description emerged from his perception of system's notes flow control. He found the sound feedback and latency between gesture and sound synthesis in a good level but stressed that the note sequence is "boring", suggesting to apply effects, such as pitch bending, to complete this gap. His main concern was about visual feedback, naming it as the priority improvement the system could have. According to him, the influence of the space in the system would reflect in a greater freedom of actions, as "running and jumping". The tiredness to use the system was also highlighted by him as a disadvantage, ironically saying that "I'm not athlete yet" for using the instrument.

3.1.3 User 3

User 3 described himself as a music lover. He used to play instruments years ago. According to him, the system was "easy to interact, but hard to play". He also complained a lot about the gestural mapping applied, saying it was not "quite intuitive" and "hard to reproduce". At most of the time, he tried to use the system exactly like a traditional instrument, trying to reproduce known songs and different musical styles. The fact that he did not reach his objective, due to the lack of feedback ("where are the [guitar] frets?") made him frustrated: "I can't even play Smoke on the Water!". Besides, he also missed the tangibility. He did not feel any kind of physical discomfort, possibly due to his athleticism, and considered as promising the potential of the system for a corporal expression context.

3.1.4 User 4

User 4 has an intimate relation with and described himself as having a certain musical background and superficial bass playing skills. He said he felt locked inside the restricted sonority and musicality of the system because he could not individually control the note execution. He suggested to use effects (bending, flange, delay) and described the system as a "simple triggering machine". For him, the visual feedback was only important to see the range of sensor's perception and mentioned the absence of a haptic feedback ("there is no physical thing to touch") as a disadvantage. He said that to use the system, the user must have "a special body control", what would be suitable for professional dancers. Besides, he said the system overloads the body, since the tiredness is evident after using it.

3.2 Examining context

All the users agreed there was practically no latency between the gesture and the sound generation. However, despite the different level of musical background of the users that could lead to different musical perceptions, the sound produced by the system was considered boring. Besides, it was a consensus that the lack of individual note control and the only manipulation of note flow parameters are not the best approach for this musical instrument, because it makes difficult to express what the user really wants. The musicality constraint was considered as a disadvantage and a suggested solution for that was the use of sound effects.

Although a clear objective, like the performance, stimulated the user to overcome the physical disturbance, the tiredness was a recurrent aspect mentioned by users. As a suggestion, the haptic feedback was mentioned as a solution to reduce the tiredness and to obtain a better guidance from the system about possible positions.

The absence of a good visual feedback was cited as a weakness of the system and the space was cited as having a close relation to the interaction freedom.

3.3 Dimension Spaces

Based on the results showed in the last section, the dimension space that summarizes subjectively the DMI here proposed was built, as can be seen in Figure 5.
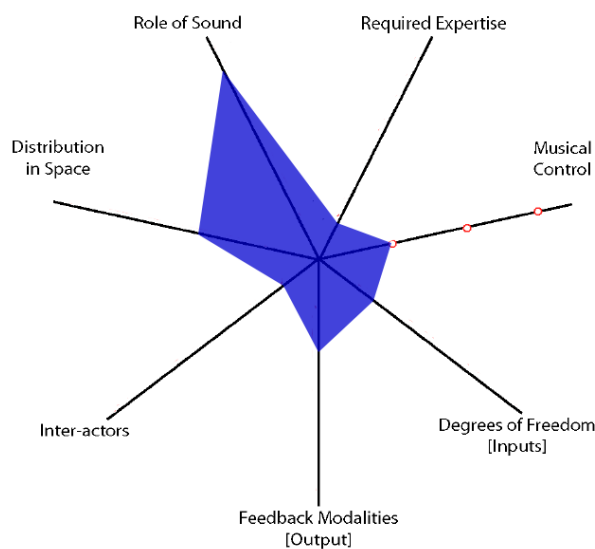
Figure 5: The 7-Axis Dimension Space of the system created based on the stage of examining context of discourse analysis

It is important to remark once again that the applied interviews were structured regarding each axis of the graph, in order to facilitate the process of plotting it. In this way, it can be rebuilt at any moment by anyone interested in the application of DA using the original material. In fact, this action is encouraged because thus there is the possibility of replicating results for better comparison, characteristic always present in effective evaluation methods.

It is also important to highlight that the exact axis-point in which the interviews were plotted is not too important, mainly due to its subjectivity. They are more meaningful when compared with others graphs of different DMIs (e.g. different prototypes of the same DMI), what provides us a simpler and cleaner way to analyze and compare them.

The graph was plotted as described, based on the stage of examining context:

a. Required expertise - represents the low level of expertise needed to interact with the system due to direct mapping between gesture and sound. It is important to differentiate this aspect from the difficulty of the proposed instrument to become a virtuoso, fact also pointed by the users; indeed, this concerns the musical control axis;

b. Musical control - the note flow control can be described as part of the "control over a musical process" category. However, it should point the problem described above;

c. Degrees of freedom - due to constraints and the restricted quantity of gestures chosen for the prototype, it was considered that the system has few degrees of freedom;

d. Feedback modalities - because of the low level of visual feedback and absence of haptic feedback, but on the other hand a good sound feedback regarding the fast response to users gestures, the point was plotted in the first third of the axis;

e. Inter-actors - as the system allows the interaction of only one person each time, the point was plotted closer to the origin;

f. Distribution in space - due to locality (to the contrary from distribution) and size of the space required, as well as the opinion of the users about it (big enough to provide certain freedom of movement, including walking and members moving), the point was placed close to the middle;

g. Role of sound - the role of sound of the system can be considered expressive, as the users considered the sound response as musical, though boring.

4 Discussion

The whole experience was considered very rich and successful, because in the end the main motivation for its application was achieved: (1) collecting user feedback for improving the development of the prototype and (2) defining a structured process that could provide more reliable analysis and comparison with other systems.

The combination of different ideas, methods and techniques has enriched the evaluation process as a whole providing more complete qualitative results and tools for comparison. However, some general considerations should be made about each step, aiming its improvement for future works.

The free exploration had an unexpected positive result during our evaluation. When the user does not have the full information about how the system works, he starts to imagine controls that might not be present in the prototype implementation. The analysis of those imaginative interactions provides rich and potential feedback results to future system interface design.

Although having a productive experience using DA, it is important to stress the amount of potential information lost when using only transcribed data, due to the complexity of human communication (e.g. facial and body expression, looks and gestures). An alternative approach that could be considered in future trials is to incorporate the process of video analysis techniques.

Another important point is the use of concept maps or mind maps in the process of itemization, which could give a powerful look to represent a preliminary version of user world and consequently an easy way to extract information and build the user world representation.

The idea to incorporate the concepts of rehearsal and performances in the solo session and group session, respectively, had a sensitive positive interference on the results of the sessions. Showing some results to an audience drove the participants concerned about having a better understanding of the system, consequently exploring it in a deeper way and amplifying the interaction experience.

Despite the dimension space proposed by Birnbaum [4] has an intrinsically quantitative characteristic (although having some subjective axis), the use of the axes based graph provides good evaluation criteria covering important aspects of a DMI making it useful to feed a qualitative approach with its parameters.

Finally, it is necessary to have in mind that this work is only a step in the process of evaluating DMIs. More researches in this area are still necessary, mainly in respect to others views, like the audience's one.

5 Conclusion

This work presented a more complete qualitative approach for evaluating DMIs considering the performer's view, based on the combination of ideas and methods of previous attempts, which enriched the evaluation process as a whole providing more complete and deeper results. It also provided a case study aiming to verify how suitable this methodology is for the context.

In order to accomplish that, the idea of using solo sessions and group sessions, with later application of discourse analysis, from Stowell et al. [2] was merged with the concepts of rehearsal and performance from W. Hsu and M. Sosnick [22]. The results were then used in Birnbaum's [4] dimension space analysis, which provided an information visualization tool for a more effective comparison with other systems.

Once some of them were not fully tested through practical experiments, this work also contributes by experimenting, confirming their effectiveness and by chaining them in a single and unified process.

Despite the good results, it is important to highlight that this method is still under continuous development, mainly due to the lack of consolidated previous works in literature. Thereby, other case studies are necessary for further comparison and checking if this alternative is indeed suitable. Another future work is improving the prototype here proposed based on the user feedback received during the evaluation process.

Finally, it is necessary to highlight the importance of this evaluation process as a phase in the cycle of user-centered design of a DMI, where user feedback is constantly used to improve the system.

6 References

[2] D. Stowell, A. Robertson, N. Bryan-Kinns, and M.D. Plumbley, "Evaluation of live human–computer music-making: Quantitative and qualitative approaches," *International Journal of Human-Computer Studies*, vol. 67, Nov. 2009, pp. 960-975.

[3] S. O'Modhrain, "A framework for the evaluation of digital musical instruments," *Computer Music Journal*, vol. 35, 2011, p. 28–42.

[4] D. Birnbaum, R. Fiebrink, J.W. Malloch, and M.M. Wanderley, "Towards a dimension space for musical devices," Proceedings of the 2005 conference on New interfaces for musical expression, National University of Singapore, 2005, p. 192–195.

[5] Preece, J., Rogers, Y., Sharp, H., 2004. Interaction design. Wiley.

[6] R.K. Merton, M. Fiske, and P.L. Kendall, The focused interview : a manual of problems and procedures, Free Press, 1956.

[7] U. Flick, An introduction to qualitative research, Sage, 2009.

[8] B. Paltridge, Discourse Analysis: An Introduction, Continuum, 2007.

[11] J. Malloch, D. Birnbaum, E. Sinyor, and M.M. Wanderley, "Towards a New Conceptual Framework for Digital Musical Instruments," Proceedings of the 9th International Conference on Digital Audio Effects, 2006, p. 49–52.

[12] M.T. Marshall, "Physical interface design for digital musical instruments," Ph.D. Thesis. MCGILL UNIVERSITY, 2010.

[13] M.M. Wanderley, "Instrumentos Musicais Digitais: Gestos, Sensores e Interfaces," *Em Busca da Mente Musical*, Editora da Universidade Federal do Paraná, 2006.

[16] N. Orio, I. Stravinsky, N. Schnell, and M.M. Wanderley, "Input Devices for Musical Expression: Borrowing Tools from HCI," Real-Time Systems, 2002.

[17] Miranda, E. R. and Wanderley, M. M. (2006). New Digital Musical Instruments: Control and Interaction beyond the Keyboard. A-R Editions.

[19] Birnbaum, D. (2007). Musical vibrotactile feedback. Master's thesis, McGill University.

[21] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-Time Human Pose Recognition in Parts from Single Depth Images," IEEE Computer Vision and Pattern Recognition, 2011.

[22] W. Hsu and M. Sosnick, "Evaluating interactive music systems: An HCI approach," Proceedings of New Interfaces for Musical Expression (NIME), 2009, pp. 25-28.

[23] Microsoft Xbox Kinect Website. http://www.xbox.com/kinect. Acessed in 1º June, 2011.

[25] Re.Scallax project website. http://code.google.com/p/rescalax/. Accessed in 1º June, 2011.

[26] SimpleSynth application website. http://notahat.com/simplesynth. Accessed in 1º June, 2011.

[27] Website containing files mentioned in this work. http://www.cin.ufpe.br/~fcac/sbcm2011. Accessed in 1º June, 2011.

[28] L.E. Wood, "Semi-structured interviewing for user-centered design," interactions, vol. 4, 1997, p. 48–61.