Beyond Evaluation: Linking Practice and Theory in New Musical Interface Design

Andrew Johnston
Creativity and Cognition Studios
School of Software
University of Technology, Sydney
andrew.johnston@uts.edu.au

ABSTRACT

This paper presents an approach to practice-based research in new musical instrument design. At a high level, the process involves drawing on relevant theories and aesthetic approaches to design new instruments, attempting to identify relevant applied design criteria, and then examining the experiences of performers who use the instruments with particular reference to these criteria. Outcomes of this process include new instruments, theories relating to musician-instrument interaction and a set of design criteria informed by practice and research.

Keywords

practice-based research, evaluation, Human-Computer Interaction, research methods, user studies

1. INTRODUCTION

As its name suggests, the focus of the New Interfaces for Musical Expression (NIME) conference is the development of new musical devices for use in live performance. Thus, a large proportion of NIME papers describe musical interfaces or instruments which show some degree of technical or artistic novelty.

The question of how to evaluate our designs has been a recurring issue. In this paper I present a framework for practice-based research in this area, in the hope that others who pursue similar work will find it of practical benefit. I argue that the process of 'evaluating' new instruments should not be seen as purely an exercise in assessment, but rather as a broader study into performers *and their creative practice* in the context of their use of the new instrument.

1.1 Evaluation and Human-Computer Interaction

Several authors have recognised the potential of human-computer interaction (HCI) techniques to investigate the experiences of performers who use musical interfaces. In general, the approach has been to use quantitative techniques from HCI which tend to equate interface effectiveness with efficiency. Wanderley and Orio [15], for example, propose a series of "musical tasks" which might be used in order to evaluate how effectively an input device can support expressive performance. These tasks are intended to create a kind of benchmark which will make it easier to compare one interface device with another. The intention is that these benchmark figures, derived as they are from formal studies of users doing prescribed musical tasks, might complement traditional technical measures of device capabilities such as output rate and precision.

This is certainly worthwhile. However, this approach is very much focussed on the devices and their ability to efficiently translate the intentions of the user into parameters for the computer. The experiences of the users who use the devices, being harder to quantify, are comparatively neglected.

To address this, we need to broaden the scope of what constitutes 'evaluation' in this context, and acknowledge that while ergonomics and efficiency are important, they are not the primary determinants of the quality of a musical interface. This thinking is reflected in the broader field of HCI, where there has been recognition that the task-based approach alone is inadequate, particularly when considering software intended to support creative work. A number of HCI researchers therefore have turned their attention to the 'user experience' [1, 10].

In addition, some researchers are proposing new ways of thinking about 'evaluation' in the context of systems which have uses that are open to a range of interpretations. Sengers and Gaver [14], for example, argue that interaction designers are becoming less concerned with designing software which unambiguously conveys and supports a clearly defined 'purpose'. They propose that HCI needs to support interactions in which users may have multiple interpretations of what a system is for and how it works. 'Evaluation' in this context goes beyond identifying whether users' interpretations of a system's purpose and behaviour matches the designer's anticipated interpretation. Rather,

> "evaluation shifts from determining whether an authoritative interpretation was successfully communicated to identifying, coordinating, stimulating, and analyzing processes of (evaluative) interpretation in practice" [14], p. 105

This approach suggests we move beyond 'evaluating' our interface designs, and use examination of users' experiences to support reflection on both musical interface design and the nature of the activities they afford. That is, we move beyond evaluating how effective our designs are at supporting musical expression and instead use them as provocative prototypes [12] which stimulate examination of the nature of expression itself – at least as it occurs in a particular cultural context.

Given this significant broadening of scope, it is timely to consider whether the term 'evaluation' is still appropriate. In my view, evaluation is best seen as a *component* of a broader examination of both musical interface design and

NIME'11, 30 May–1 June 2011, Oslo, Norway.

musical expression and I therefore argue that a more general term such as 'user experience study' is preferable. It is certainly important that we evaluate our instruments - that we assess how well they meet relevant criteria - but because our design criteria embody our theories of designing for musical expression, we should be equally interested in refining, or redefining, the criteria.

2. RESEARCH STRUCTURE

The research process I have adopted is shown in figure 1. Initial design criteria, drawn from the literature and personal experience inform the design of new musical interfaces (or instruments for want of a better term). From the design process we get the musical interfaces themselves and a set of design criteria which the designer believes they embody. These instruments, and the experiences of musicians who use them, are scrutinised in a series of user studies. From these studies we gain theoretical understanding of musical performance with instruments of this kind and a refined set of design criteria informed by practice and research.

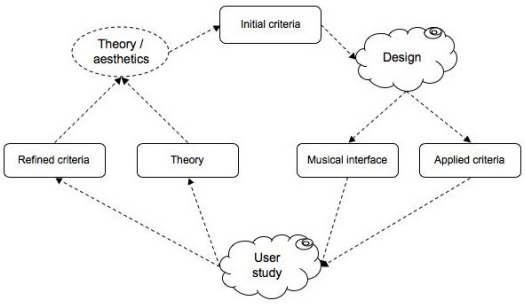

Figure 1: Overview of the research process.

I will consider this process in some detail in the following sub-sections.

3. DESIGN CRITERIA

When presenting new musical instruments, it is possible to describe the instruments in purely functional terms - to outline exactly what they do - but it is equally important to consider *why* they work in this way. Of course, the instruments behave the way they do because the designers made them that way, with particular goals in mind. Identifying these goals - the design criteria - helps make the intentions of the designers explicit and therefore open to examination and discussion. It also facilitates evaluation of the new instruments in terms of the criteria.

Portillo and Dohr define a design criterion as, "a measure of value used by the designer to conceptualize, test and evaluate the project purpose in the design process" [13], p.405. An important point to note here is that design criteria are used, perhaps tacitly, during the design process as the designers develop their ideas and the designed artefacts in parallel. Making these criteria explicit may not be easy, and producing an exhaustively complete list of all criteria that were applied is impractical and probably not particularly helpful. A balance and appropriate level of granularity needs to be found.

In my research I have drawn on several sources to identify initial sets of design criteria that were most influential in shaping the instruments I have created. These include:

- Reflective online diaries (blogs) kept by those involved in the design process (composers, musicians, software developers, etc).
- Interviews with artists and designers involved in the development of the instruments.
- Examination of software version control logs.[1]

It is important to note that it is not expected that all the applied design criteria should, or could, be identified at this stage. Criteria identified prior to the user studies help focus the interviews and user studies which follow. We expect, and hope, that criteria will be significantly refined and added to as the research progresses.

4. USER EXPERIENCE STUDIES

User studies are primarily concerned with three questions:

1. Do the instruments that have been created meet the design criteria identified during design?
2. How do musicians experience them?
3. What are the relationships between the characteristics of the instruments and the musicians' experiences?

I have conducted studies of a series of instruments [6, 8, 9], and in this section I briefly outline the approach to data gathering and analysis. Data was gathered from the following sources:

- A user study in which seven professional musicians were videoed playing with the instruments, commenting on their experiences and responding to interview questions. This was the primary source of data.
- Notes made by observers who attended the musicians' sessions with the instruments. These provided additional perspective on the musicians' experiences and helped identify whether the instruments met the design criteria.
- Questionnaires administered during the musicians' sessions with the instruments which attempted to directly elicit their opinion on whether it met the design criteria.

The question of whether the instruments met the design criteria was primarily addressed by analysing the questionnaires and the notes made by the observer. The more complex question of how musicians interacted with the instruments was the primary focus of the study. In order to address this question, the video recordings of the musicians' sessions with the virtual instruments were transcribed and the grounded theory method was used to generate a theory of musician - interface interaction. The observers' notes provided additional perspective on this data and informed the development of the theory.

The studies involved seven professional musicians who had a minimum of five years professional experience. They included principal players from professional symphony orchestras as well as leading improvisers. Due to the degree of expertise of the participants and the in-depth nature of the evaluation, this was a sufficient number to provide detailed insight into the experiences of expert musicians with the virtual instruments. Note that in qualitative research

[1]Version control software (eg. Subversion) is used during software development to track changes at every stage of the design process.

the emphasis is on *generating*, rather than validating, theory [5]. As such, this research was intended to provide detailed insight into the experiences of the specific musicians who participated in the evaluation, and to generate theories consistent with what was observed. It is hoped that this research will provide a sound basis for future research which may attempt to more broadly *validate* the concepts and relationships uncovered in this study. Such validation would be likely to involve larger numbers of musicians using virtual instruments in a simplified and more controlled context.

The focus of the investigation was on what the musicians were able to do with the virtual instruments, what impact using them had on their music making and any suggestions for improvements, so the musicians were not asked to perform specific musical tasks. Rather, they were told in simple terms how the virtual instruments behaved and then asked to explore and make music with them.

The musicians were asked to verbally reflect on their experience with the instruments using a variation of the 'think aloud' approach [3]. When using the concurrent think-aloud approach, the idea is that the musicians continuously verbalise what is going through their mind as they use the instruments, keeping the time between thought and verbal expression to a minimum. However, asking musicians to generate fully concurrent think-aloud reports presented obvious practical problems because wind and brass musicians are unable to speak (intelligibly) and play their instrument at the same time. A sensible compromise was to ask the musicians to verbally report what they were thinking and perceiving as frequently as they were able during their time using the instruments. This meant that they were effectively providing a large number of smaller retrospective reports as they played for a time, commented on what was happening, played some more, made further comments and so on.

In addition to gathering information about what the musician was thinking and experiencing as they used the virtual instruments, the musician's opinions on the instruments and suggestions for how they could be improved were actively solicited. As experts in their field, it was hoped that the musicians would be able to provide insight into the nature of the virtual instruments, their potential uses, limitations and areas for improvement. The intention was that the musicians would become engaged with the design process and in a sense become co-designers. As such, the format of the evaluation was flexible. There was a standard procedure but when interesting issues arose, this was varied. Because the emphasis of this study was on theory generation rather than verification, the gathering of rich data was prioritised over consistency of procedure. The process was more akin to a user dialogue than usability testing [2].

After using each virtual instrument, a semi-structured interview was conducted in which participants were asked a series of open questions relating to their experience with the virtual instrument. In order to facilitate later analysis, the musicians' interaction with the instruments and the interviews were video recorded.

4.1 Data Analysis

The video-recordings of the musicians playing the virtual instruments and talking about their experiences were a very rich source of data. A challenge was to identify consistent themes and patterns in order to make sense of this information. Techniques from the grounded theory method [5, 4] were therefore used to code and analyse the data gathered. This method was a good fit for this purpose because it facilitated the generation of theory closely tied to the evidence from rich qualitative data. At a high level, the basic steps of the grounded theory analysis process as applied in this study were:

- Transcribing the evaluation sessions.
- Open coding: that is, identifying and labelling incidents in the data (including non-verbal data). This is done line by line, coding each sentence. As coding progresses, incidents are constantly compared with one another to identify similarities and differences.
- Memoing: as ideas emerge regarding the codes and their relationships during coding, the researcher stops to make a note. Memoing aids the process of linking the descriptive codes into theory.
- Sorting: memos are sorted and arranged in order to identify core issues and their relationships with one another and thus build theory which is 'grounded' in the gathered qualitative data.

In my work I have made use of the open-source software Transana [16] to facilitate this process. With Transana, clips of interesting video data can be created and labelled with codes (known as 'keywords in Transana) which are specified by the researcher. Once coding is complete, searches can be made which find all clips assigned particular codes. For example, a search could be made which found all video clips from all participants which were assigned the code 'control'. Each of these clips could be examined in detail to find key points of similarity and difference. These features were invaluable when dealing with the more than fourteen hours of video gathered during the studies.

4.2 Building Theory

Obviously, merely labelling incidents in the data does not create theory, but building up a coding scheme in this way facilitates what Glaser and Strauss [5] describe as the 'constant comparison' technique. Constant comparison simply involves comparing incidents in the data with one another, identifying similarities, differences and relationships which are recorded in memos as the researcher identifies them. In the grounded theory method, memoing is the process by which the analyst reflects upon and documents their evolving understanding of the situation under study. Memoing also helps the analyst to link the codes together into a theoretical framework. Memos are simply notes written by the researcher. They do not have a required format, the intention being simply that insights are captured quickly so that they are retained. Memoing in this study made use of a feature of Transana which allows the researcher to attach 'notes' to transcripts or collections of clips.

Through this process the researcher builds a theory which helps to make sense of the situation under examination. Memos help facilitate and, to some extent, document the researcher's evolving understanding of the links between these incidents. However, it is important to note that memos and coding schemes are not a complete record of the analysis process. In my research, coding and memo-writing are undertaken primarily to facilitate analysis rather than document it. Thus the coding scheme and memos should be considered a by-product of the analysis process which generates theory.

The fundamental idea is that the researcher examines the codes that have been created during open coding and attempts to identify higher-level concepts that make apparent patterns in the codes, and relationships between them. The approach described above draws primarily on the suggestions of Glaser [4] and Miles and Huberman [11].

4.3 Findings

This paper is primarily concerned with research methods and space precludes a detailed discussion of findings which have been published elsewhere [6, 7, 8]. However, I will briefly outline some key findings in order to illustrate the kinds of conclusions that can be drawn from a study of this type.

During grounded theory analysis it is expected that open coding will lead to the discovery of a 'core' category, a key issue which appears to have particular relevance to the situation under study [5]. The core category emerges during analysis as the researcher continually compares incidents in the data, noting relationships between incidents in memos.

When analysing the data gathered during this study, it was clear that the musicians did not always approach the virtual instruments in the same way. Sometimes a musician would express frustration because they felt they did not have enough control over the virtual instruments, but then at other times the same musician would complain that the virtual instruments were not autonomous enough, and that they wanted their behaviour to be less predictable. It seemed that the qualities the musicians sought in a virtual instrument would change during their interactions - that they interacted with the virtual instruments in different modes. Thus the core issue which emerged during analysis was that of *modes of interaction*.

We found that the musicians interactions with the virtual instruments could be classified into three modes: instrumental, ornamental and conversational. In instrumental mode the musician seeks a high level of detailed control over all aspects of the virtual instrument's behaviour. Musicians taking an instrumental approach essentially see the virtual instrument as an extension of their acoustic instrument and want it to respond consistently so that they can trust it during performances.

In ornamental mode, musicians surrender detailed control of the generated sound and visuals and let the virtual instrument create audio-visual layers that are added to their acoustic sounds. Musicians taking an ornamental approach may not pay active attention to the behaviour of the virtual instrument, instead leaving it to its own devices and expecting (or hoping) that it will do something that complements or augments their sound without requiring directed manipulation.

Conversational interaction occurs when musician approaches the virtual instrument as a musical partner. In conversational interaction the musician allows the virtual instrument to 'talk back', at times directly influencing the overall direction of the music. The musical 'balance of power' is in flux as responsibility for shaping musical direction continually shifts between musician and virtual instrument.

5. CONCLUSION

In this paper I have detailed an approach to linking practice and theory in musical interface design. The guiding principles of this method have been described and I have summarised how it has been applied to generate and refine theory concerning the nature of performers' interactions with musical interfaces.

The outcomes of the practice-based research process I have outlined are a set of musical interfaces, a theory of musician-instrument interaction and a set of design criteria informed by practice and research.

I believe that criteria-based evaluation and qualitative user studies are a simple, yet powerful combination which enables a form of detailed and rigourous reflection on the creative outcomes of musical interface design. The specific methodological choices I have made in relation to how to gather and analyse data were driven by the particular characteristics of the musical interfaces we designed and the aesthetic goals which guided their development. Thus, I do not propose this method as a detailed one-size-fits-all solution, but hope that discussion of this work will encourage a broader view of 'evaluation' in musical interface design and help practitioners and researchers more effectively link practice and theory.

6. REFERENCES

[1] M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright. *Funology: from usability to enjoyment.* Kluwer Academic Publishers, Norwell, MA, USA, 2004.

[2] J. Buur and K. Bagger. Replacing usability testing with user dialogue. *Communications of the ACM*, 42(5):63–66, 1999.

[3] K. A. Ericsson and H. A. Simon. *Protocol Analysis: Verbal Reports as Data.* MIT Press, Cambridge, MA, revised edition, 1993.

[4] B. G. Glaser. *Theoretical Sensitivity.* The Sociology Press, 1978.

[5] B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: strategies for qualitative research.* Aldine de Gruyter, New York, 1967.

[6] A. Johnston. *Interfaces for Musical Expression Based on Simulated Physical Models.* PhD thesis, University of Technology Sydney, 2009.

[7] A. Johnston, L. Candy, and E. Edmonds. Designing and evaluating virtual musical instruments: facilitating conversational user interaction. *Design Studies*, 29(6):556–571, 2008.

[8] A. Johnston, L. Candy, and E. Edmonds. Designing for conversational interaction. In *Proceedings of New Interfaces for Musical Expression (NIME)*, 2009.

[9] A. Johnston, B. Marks, and L. Candy. Sound controlled musical instruments based on physical models. In *Proceedings of the 2007 International Computer Music Conference*, pages vol1: 232–239, 2007.

[10] J. McCarthy and P. Wright. *Technology as Experience.* The MIT Press, 2007.

[11] M. B. Miles and A. M. Huberman. *Qualitative Data Analysis: An Expanded Sourcebook(2nd Edition).* Sage Publications, Inc, 1994.

[12] P. Mogensen. Towards a provotyping approach in systems development. *Scandinavian Journal of Information Systems*, 4:31–53, 1992.

[13] M. Portillo and J. H. Dohr. Bridging process and structure through criteria. *Design Studies*, 15(4):403–416, 1994.

[14] P. Sengers and B. Gaver. Staying open to interpretation: engaging multiple meanings in design and evaluation. In *DIS '06: Proceedings of the 6th conference on Designing Interactive systems*, pages 99–108, New York, NY, USA, 2006. ACM.

[15] M. M. Wanderley and N. Orio. Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal*, 26(3):62–76, 2002.

[16] D. Woods and C. Fassnacht. Transana v2.22. Madison, WI: The Board of Regents of the University of Wisconsin, 2007.