The Show Must Go Wrong: Towards an understanding of audience perception of error in digital musical instrument performance

S. M. Astrid Bin

PhD thesis

Media and Arts Technology PhD Programme
School of Electronic Engineering and Computer Science
Queen Mary University of London

2018

Abstract

> *... the most interesting thing in terms of art would be to infiltrate the spongy encephalon of the modern viewer. Because the mystery now resides there, in the viewer's brain ... what is its secret?*
>
> Jean Baudrillard, *Art ... Contemporary of itself*, 2003 [16, p. 93-94]

This thesis is about DMI (digital musical instrument) performance, its audiences, and their perception of error.

The goal of this research is to improve current understanding of how audiences perceive DMI performance, where performers and their audiences often have no shared, external frame of reference with which to judge the musical output. Further complicating this audience-performer relationship are human-computer interaction (HCI) issues arising from the use of a computer as a musical instrument. In current DMI literature, there is little direct inquiry of audience perception on these issues.

Error is an aspect of this kind of audience perception. Error, a condition reached by stepping out of bounds, appears at first to be a simple binary quantity, but the location and nature of those boundaries change with context. With deviation the locus of style and artistic progress, understanding how audiences perceive error has the potential to lend important insight to the cultural mechanics of DMI performance.

In this thesis I describe the process of investigating audience perception and unpacking these issues through three studies. Each study examines the relative effects of various factors on audience perception — instrument familiarity and musical style, gesture size, and visible risk — using a novel methodology combining real-time data collected by mobile phone, and post-hoc data in the form of written surveys. The results have implications

for DMI and HCI researchers as well as DMI performers and composers, and contribute insights on these confounding factors from the audience's perspective as well as important insights on audience perception of error in this context. Further, through this thesis I contribute a practical method and tool that can be used to continue this audience-focused work in the future.

Acknowledgements

Without the still and deep support of my parents who always just kind of let me get on with it no matter how ridiculous the project, this work would not have been possible. Without Pete's patience, forbearance, sense of humour, technical skill, critique, and ability to cook meals and take out the rubbish, it would also not have been possible. Mum, Dad, Pete: Thank you for everything.

I simply could not have imagined better supervisors. Dr Andrew McPherson and Dr Nick Bryan-Kinns have been supportive, challenging, generous, corrective, flexible, and helpful at every stage of this work. I am grateful to have had permission to bother academics of such quality for four years.

Thank you to my colleagues in the Media and Arts Technology program, the most inspiring, clever, funny and capable bunch of anarchists to have ever crash-landed in an engineering department. Thank you to Liam Donovan, Daniel Gabana, Julie Freeman and Vanessa Pope especially, for their creativity, skill, sense of humour, and willingness to make it work.

Without the collaboration, patience, generosity, and forgiveness of the following QMUL staff members I would have accomplished nothing: Jules Deering, Hazel Coville, Hayley Cork, Sharon Cording, Julie Macdonald, Ho Huen, Dave Wilkinson, Jonathan Winfield, Karen Bray, Geetha Bommereddy, EECS technical support, Kelly Peake and the Thinking Writing team. Thanks also to MAT's original anarchist Richard Kelly, whose enthusiasm was the main reason I did this degree in the first place.

Thanks additionally to: Saints Choronzon, Head Crash and Grimmware for favours received; Fabio Morreale for his good advice; Marcus and Joanna; the 2600 crew; The Tower; Leila and Sophie; 24/7 access to the workshop; the plains of Leng; Dr Graham White for the hallway conversations/oblique impacts; Megabyte for annoying me every night at midnight so I would go to bed. Thank you to Ian Carr-Harris, David Macintosh and Maria McCarthy, whose influence has cast a long, long shadow. Thank you also to those who kept my love of music and my desire to say something about it burning all these years, most especially Reverend Mark, P. Lankton, Bill Drummond, Bobby Richardson, and every 60Hz hum that has in turn kept my own motor ticking over.

Thank you also to the brilliant musicians who have taken part in this research, without whom none of it would have happened: Robin Rimbaud, aka Scanner; Dianne Verdonk; Tim Exile; Benjamin Duvall, Ben Fair and Jon Hering of Ex-Easter Island Head; Joanne Armitage; Laurel Pardue; Bex Burch; Zands Duggan; Rosie Bergonzi; Callie Hough; Henrik Langer.

Thank you, finally, to the EPSRC for funding a centre for doctoral training that has benefited artists like me. By completing this PhD I have gained technical skill that enabled me to make things that I previously could not have imagined. The skills I have honed and expanded through this process have certainly made me a better technologist, but more importantly I am emerging from this experience a better artist. May many more artists be so fortunate.

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) as part of the Doctoral Training Centre in Media and Arts Technology at Queen Mary University of London (ref: EP/G03723X/1).

Associated publications

Portions of the work detailed in this thesis have been presented in national and international scholarly publications, as follows:

- Chapter 5: Bin, S. M. A; Bryan-Kinns, N,; McPherson, A. P. *Skip the pre-concert demo: How technical familiarity and musical style affect audience response.* New Interfaces for Musical Expression 2016; Brisbane Australia.

- Chapters 3 and 5: Bin, S. M. A.; Morreale, F.; Bryan-Kinns, N.; McPherson, A. P. *In-the-moment and beyond: Combining post-hoc and real-time data for the study of audience perception of electronic music performance.* ACM Interact 2017; Mumbai India.

- Chapter 7: Bin, S. M. A.; Bryan-Kinns, N,; McPherson, A. P. *Hands where we can see them! Investigating the impact of gesture size on audience perception.* International Computer Music Conference 2017; Shanghai China.

- Chapter 8: Bin, S. M. A.; Bryan-Kinns, N.; McPherson, A. P. *Risky business: Disfluency as a design strategy.* New Interfaces for Musical Expression 2018; Blacksburg Virginia USA.

‘To the makers of music — all worlds, all times.’
Voyager 2

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is about audience perception of digital musical instrument (DMI) performance. Specifically, it examines the nature and function of error within this musical practice, and how it is perceived by audiences. Three studies, each with a live audience watching musicians in a concert setting, seek to unpack some of the confounding aspects of DMI performance that relate to error — familiarity, gesture, and risk — that are currently not well understood.

Though the audience is an essential element of musical performance, it remains understudied in the DMI research community. Using a novel methodology developed for this work, real-time and post-hoc data were collected from participating audiences, and viewed together to gain insight into how audience members perceive DMI performance, how they perceive error, and the role error has to play in this musical practice.

Since there is little existing literature that explores how audiences perceive DMI performance, this thesis takes an exploratory stance. It aims not to construct fixed models of audience understanding, but rather seeks insight into the relevant audience-related questions. Through this process this work also illuminates and identifies potential pathways for future research.

1.1 Motivation

DMI performance is defined by its radical technology-led experimentation, that applies not only to the instruments used but the music produced (Chapter 2 begins with an examination of how this performance practice came to reject Western classical musical vernacular). Though this experimentation without limits is an exciting creative opportunity, it has had a curious knock-on effect: In a wide-open space of limitless possibility, with no shared frame of reference to guide them, how can audiences know what a 'successful' performance sounds like? Inversely, how can audiences possibly detect 'error'?

Error, on its surface, is a simple concept. From the Latin *errare*, or 'to stray', it denotes a deviation, an action out of bounds, and presents as a straightforward and binary quality. But error quickly becomes complicated when we examine the nature of those boundaries. Who puts them in place? What are they demarcating? What function do they serve? What is at risk if they are overstepped? Is overstepping always a bad thing? Who decides?

The location and boundaries that define error change with context, as well as the impact and perception of an error's outcomes. Some errors are catastrophic; an error committed by a pilot or a surgeon could have fatal consequences. Within creative domains, however, deviating from the norm is often the locus of progress, accidental discovery, and creative opportunity. In the Western classical music tradition there is enormous value placed on accuracy in the reproduction of score and composer intent, and exacting guidelines on how musical performance should be executed (with made for 'interpretation' as it applies to tempo and phrasing, which in turn are governed by stylistic norms). With a high value placed on playing technique that accurately reproduces the score, error is obvious, and is seldom seen as anything except something to be mitigated and against which a player is insured through exhaustive practice.

DMI performance, by contrast, has no such aesthetic goals. DMI performance as it exists today arose out of early 20th century avant-garde through experimental electronic pioneers (including Pierre Schaeffer, Iannis Xenakis, Daphne Oram, John Cage, and various others) who valued radical, technology-led experimentation, and rejected hegemonic influences of score, stylistic expectations and aesthetic constraints in favour of experimentation and newness. This performance tradition has continued and developed, and since the 1990s — when personal computers became small and powerful enough to be useful to electronic musicians — DMI performance has centred around the computer.

In this thesis I argue that this lack of score and cohesive playing tradition mean that the boundaries essential for error to exist become hazy.

Compounding this murkiness are the confounding factors that have come along with the computer as sound-producing device. The most prominent, termed by Miranda and Wanderley as *control dislocation* [141], means that the sound that is made by these instruments can be entirely unrelated to the materials they are made from and the way they are played. This breaks with 30,000 years of human musical tradition [175], where the sound an instrument makes is intrinsically connected to its materiality and manipulation.

Running parallel to the DMI performance community is the DMI research field. This vibrant field of academic inquiry, mainly located in the New Interfaces for Musical Expression (NIME) community, is where scholarly research on how the D in a DMI affects the music that we can make. The people doing DMI research are often also DMI performers, and the research and the artistic output tend to inform one another.

DMI research is an active field of research on sound-production techniques, performance, instrument design, and so on, but curiously the audience — an essential element in musical performance — remains largely unstudied, and focus has instead rested on the experience of the performer. In the preliminary research that exists on error in the DMI research field on the audience experience of error [74, 75, 75], one study suggested that, perhaps, *error was not even possible* in DMI performance.

This is an intriguing conclusion that raises more pertinent questions. DMI performance makes music that is often radically experimental and for which there is no frame of reference, with instruments that make sound that's unrelated to the materials they are made from or the way the performer manipulates them. This is certainly poses a unique challenge for audiences, and may explain why, in the words of Bob Ostertag, DMI performance 'can find no audience beyond those who create it' [155]. Though commercial success or popular taste are anything but suitable arbiters of artistic quality, this is an intriguing phenomenon that has received little attention. It's been suggested before that a method of combating this lack of audience understanding would be to establish systems of scores and playing traditions to create this missing frame of reference [62], but this approach limits the very features that define DMI performance practice.

The motivation for this thesis, then, is this: To understand how error works in this context by studying the perception of audiences, to discover ways we might design DMIs that do not limit DMI performance's radical experimental spirit, yet create the conditions for error to exist in the minds of those who watch it.

1.2 The context of this research

The term 'DMI performance', in this thesis, refers to a type of musical practice that uses novel digital interfaces to create music that is often abstract or exploratory in its approach.

Though the NIME community has contributed considerable literature exploring the technical aspects of DMIs, as well as reflections on working methods and creative goals of this artistic practice, the short history of DMI performance as it is currently understood dates back only to the 1990s with the emergence of DMI practice and the NIME community itself. As a result, comprehensive examination and definition of this art form, let alone a canon or debates around what such a canon would include, has yet to develop within or around this artistic community. There are some works that define aspects of the tools and artistic aims of DMI practice (such as [141], or more recently [31]), but there is currently a lack of literature that exists to trace the artistic goals of individual practitioners or the community at large, as well as work that draws connections between DMI performance and the cultural and socio-technical influences that have shaped it over time.

In order to establish a starting point, Chapter 2 begins with a tracing this history by drawing a line that runs from the early 20th century avant garde, through the work of Cage, continuing through the 1990s when DMIs emerged, through to the present day. In this way I present theoretical perspectives on how DMI performance's radical and technology-led experimentation has evolved and the influence of Western classical music tradition on this evolution, and the ways in which DMI performance still retains some of this tradition's formal constraints, discussed in relation to Goehr's theory of *Werktreue* [83].

Additionally, Chapter 2 includes a survey of contemporary DMI performance practice. This community is a lively area of artistic output, but neither its boundaries nor its connection to wider cultural trends are clearly defined. In this chapter I provide an overview of current DMI practice not to reduce it to a stable definition, but rather to demonstrate its breadth, diversity and flux, and to define and contextualise the type of practice to which the term 'DMI performance' refers throughout this thesis.

DMI research sprawls across several related domains that include music, engineering, design, and human-computer interaction, each deeply influential on DMI performance practice and the instruments it uses. Chapter 2 includes a survey of the the values, goals, knowledge, ways of working and communities of practice that exist in these domains, and how they inform DMI research and impact DMI performance.

Additionally, Chapter 2 surveys how these related domains each define and understand the nature of error, presenting a plurality of ideas of what error is and what it does. This provides a starting point for understanding how error might operate in DMI performance, where there is currently only the most preliminary knowledge on how audiences understand error, and where the audience is largely understudied.

This exploration of the history of and influences on the DMI domain constructs a firmer understanding of this community of research and artistic practice. Neither DMI practice nor research is simply a sum or amalgamation of these related disciplines, but rather a domain of research and artistic practice in their own right. However, untangling these separate influences and drawing connections between these and the community has the dual benefit of providing a point of departure for this research, and also contributes to understanding of the nature of the domain of DMI practice itself — one that is, currently, not cohesively documented by either its practitioners or outside observers.

There are a number of confounding factors facing audiences of DMI performance, and Chapter 2 surveys the existing knowledge and approaches around these. These include the notion of transparency from Fels et al [68], Miranda and Wanderley's notion of control dislocation [141], as well as their theories around gesture, as well as ideas about audience perception of skill and error from Fyans et al. [74, 75, 72]. This section serves to highlight the gap in knowledge around audiences, as well as the need for a new way of studying them if we are to understand how they experience error in DMI performance. Additionally, it provides an overview of the subject matter of each of the three studies in this thesis.

1.3 How this work was carried out

The studies described in this thesis use one methodological approach, developed specifically for this body of work, that combines both real-time and post-hoc audience data. This combined approach was developed to produce a nuanced and multi-faceted view of audience perception of an artistic work. An in-depth description of this methodology is in Chapter 3, including an exploration of how audiences have been studied in DMI and related research domains, and highlights the opportunity for a more nuanced way of studying audiences.

Chapter 3 also explores the rationale for carrying out this work in a live context. It explores the multimodal nature of the live music experience, and details some sources of complexity inherent in a live experience. This complexity, though it can be confounding, can also add to the veracity of the multimodal experience and lend insight that could not be captured in a lab or recorded setting. I also explain how complexity was controlled and encouraged in order to create a situation that was both a legitimate performance experience and could also be credibly studied.

Finally, Chapter 3 details my research approach and goals. As there is little established work within the DMI research field on which to draw upon or test against, this thesis is exploratory by nature. I briefly describe my post-structuralist approach that accepts that there may be a plurality of meaning and that meaning is not fixed. By extension, this work is not intended to propose formal models of audience experience — this is beyond the scope of this thesis. Instead, this thesis applies the combined methodology to gain insight into error's role in DMI performance as well as audience perception of live music, to test the impact of DMI design interventions on audience perception, and to highlight intriguing areas for future work.

Part of this novel combined methodology is a real-time data collection system called Metrix. This is described in-depth in Chapter 4, including the need for such a platform, the technical architecture, the process of interface design and iteration, techniques used for audience onboarding, markers of success and areas for further development and improvement.

1.4 The three studies

This thesis describes three studies. The overarching research question of the nature and function of error in DMI performance takes contributions from all three studies. However, there are confounding questions for which no audience research currently exists that must be unpacked in order to understand error to any degree. These confounding factors are centred around transparency, gesture and risk.

Study 1, described in Chapter 5, takes on the most fundamental of these factors: The playing style of DMI performance is often abstract and experimental, and the instruments used to play it are often completely unknown to the audience. DMI research has in the past focused on making the instrument more familiar to the audiences as a way of mitigating this, but there has been no study that separates the relative influence of the familiarity of the instrument and musical style on the audience experience. This first study unpacks precisely this question, and Chapter 5 describes the study design, method, results, and findings, using the methodology that combines the survey and real-time data in order to gain nuanced insight into audience perception of DMI performance.

Studies 2 and 3 used instruments that I designed and produced specifically for these experiments. Chapter 6 describes the rationale for undertaking this significant challenge, and details the design approaches that underpinned this process, as well as how the goals of the study were met through design. It contains a description of the approach to materials, my own design values that provided a point of departure, and contains an in-depth look at the approach not only to the physical design of these instruments, but also the design of the sensor processing system and internal hardware.

Study 2 is detailed in Chapter 7, and explores gesture, and how changing the scale of a DMI can affect audience perception. Gesture is an often-talked about term within DMI literature and there is general consensus that gesture is meaningful, communicative, and important, but there little indication of what a communicative gesture might look like. While gaining further insight on audience perception of error, this study also suggests that instruments that require more visible gesture may positively impact audience enjoyment, and explores how this is reflected in the qualitative, quantitative, and real-time data.

Study 3 was the final study of this thesis, and is found in Chapter 8. Carrying forward the findings from Studies 1 and 2, this third study explores how visible risk, created by disfluent DMI behaviour, might affect performer choices that in turn impact audience perception. For this study I produced Keppi, a percussion instrument, in six identical versions. Each had one of three disfluent behaviour states that provided the 'risk' state, and each was played by an experienced percussionist in a concert setting, and I gathered both real-time and post-hoc data from the audience participants. The results of this study did not confirm that disfluency resulted in visible risk, but this study did find that the audience notices and responds to control and effort, or visible skill. Further, the results of this study suggest that design can facilitate a plurality playing methods to allow musicians to leverage their skill and personal style — factors that I found audiences noticed and responded to, providing an intriguing avenue for further study and a useful insight for DMI designers.

1.5 Understanding the results

The outcomes of this exploratory thesis are both practical and theoretical. The theoretical insights, though explored in-depth at the end of each study chapter, are gathered together in Chapter 9, which also presents a meta-analysis of all the research outcomes of this thesis.

Chapter 9 also suggests ways that these insights can be applied by DMI designers and practitioners that will allow for compelling audience experiences while still maintaining the exploratory, radical, and technology-led spirit of DMI performance practice that have made it an artistic practice that has evolved for more than a century. Finally, I suggest ways that DMI performance practice, precisely because it is shot through with human complexity, confounding technological factors and artistic goals, is ripe ground for study of how we use digital systems for very human creative goals, and may provide insight to other domains grappling with the same questions.

This thesis has also uncovered potential avenues of future work, which are detailed in Chapter 10. This chapter includes reflections on the challenges and lessons that have emerged from studying this complex artistic practice, and lessons learned from studying live audiences in an experimental context. I also suggest ways that these insights may apply to other domains where computers are used as tools of creation, so the digital tool may serve, and not limit, the human creator.

1.6 Research questions

There are four questions guiding this research:

- **Question 1:**
  With no external stylistic frame of reference, how does the audience perceive error in DMI performance, and how does this affect enjoyment?
- **Question 2:**
  What is the impact of visible risk on audience perception of DMI performance?
- **Question 3:**
  Can the physical design of a DMI affect the performative outcomes?
- **Question 4:**
  How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

1.7 Contributions of this thesis

This thesis makes three theoretical contributions to knowledge in the space of DMI research, as well as two practical contributions.

The theoretical contributions are as follows:

- New insights into error, and how it is perceived by audiences in DMI performance.
- A historical and cultural perspective on DMI performance, and a review of current practice.
- Insights into designing DMIs for experimental use.

Along with these theoretical contributions, this thesis also makes the following practical contribution:

- A new methodology for live audience research, that combines post-hoc and real-time data, and Metrix, an open-source software tool for real-time data collection.

The following chapter, Chapter 2, presents a survey of the existing literature related to this research, casting a wide net over a range of disciplines and areas of influence. In this way, I demonstrate the need for this investigation, and identify gaps in the literature that are addressed by the above questions.

Chapter 2

Related work

The topic of this thesis — audience perception of error in digital musical instrument performance — lies across a number of disciplines. This chapter casts a wide net over a range of existing knowledge in fields that are related but often disperate.

Music is a cultural experience, and as such, it is important to understand any particular tradition of music within its social and historical context. Therefore, Section 2.1 traces the roots of DMI performance to provide a wider historical perspective, from the early pioneers of electronic music and the influence of modernism through to the late 20th century emergence of DMIs as we know them today.

There are a number of models of how audiences perceive music, and Section 2.2 surveys these. Beginning with Lydia Goehr's theory of *Werktreue*, I also draw on literature from HCI and music perception research in light of this model. I also make the important distinction between *Werktreue* and musical vernacular, using the work John Cage to illustrate this.

The computers-as-instruments model of DMIs emerged in the 1990s. The shift to thinking of a computer as a musical instrument has disrupted long-established ideas of what an instrument is, and as a result there are aspects of DMIs that are confounding for audiences. In Section 2.3 I survey these factors and the contemporaneous research around them from the DMI research domain. Additionally, I examine the influence of HCI on this research. DMI research has its roots in the HCI research community, and HCI's influence has popularised certain approaches and ways of thinking, particularly the de-emphasising of the audience in the existing literature in favour of the perspective of the performer.

Section 2.5 examines the potential of error as an area of inquiry in DMI research. Departing from existing theory in the DMI community, this section also casts a wide net across the related disciplines of music, engineering, HCI and design to present a varied and nuanced view of what error is and what it means. I also examine error in a musical context, and argue for its usefulness in light of the results of removing error from recorded music via Auto-Tune. This section proposes that error may be a locus of knowledge about audience perception of DMIs that has been previously overlooked and worthy of closer study.

Finally, Section 2.6 concludes with a summary of this chapter, and specifies the questions that have guided the three studies described by this thesis.

2.1 Digital Musical Instruments: Historical and contemporary practice

In this section, I provide a historical context for DMI performance practice by tracing the evolution of this art form from its early beginnings to the present day. Then, to provide a point of reference for the term 'DMI performance', as well as to demonstrate the breadth and diversity of this art form, I also provide a brief overview of current activity within this space by relevant practitioners. Finally, I articulate the type of DMIs that are considered in this work.

2.1.1 The influence of modernism

In a 1923 essay, Virginia Woolf famously remarked, 'On or about December 1910 human nature changed.' [210] Her words were not unfounded: In the early 20th century the existing power structures of politics, economics, art, music and technology in Europe (and, by extension of influence, North America) had begun to shake. From about 1910 through the 1950s, these changes had profound effects on all parts of society.

These dramatic changes were due to a number of influences. World War 1 from 1914-1918 had brought the horror of conflict into a civilian context, through automatic weapons, aviation, and chemical warfare. At the same time, other shifts were occurring: Mass migration into cities was taking place; work was becoming ever more mechanised; electricity was beginning to be widely available in homes; widespread literacy had started the proliferation of mass media.

These changes were reflected not only in the economic, social and political spheres, but were also deeply resonant in art and creative practices. In 1934 Ezra Pound proclaimed 'Make it new!' [165] and, by that time, the making of newness was well underway. After the devastation brought by the World War 1, many artists rejected the expectations of bourgeois culture, and instead were led by a desire to forge a new world using new technology. From architecture to literature to film, modernists were preoccupied with the symbols of progress, and refusing to look back.

2.1.2 New sounds and new instruments

Music, of course, was not immune to these shifts. Composers such as Debussy and Strauss were already experimenting with elements of musical vernacular (such as tone, form, rhythm, pitch), and more avant-garde composers were re-imagining music altogether, such as Russolo and his Futurist Manifesto [174]. Russolo was inspired by the new, mechanical soundscape of industrialisation and cities, and called for these new sounds of the mechanised world to be considered musical elements in their own right. It was soon clear, however, that entirely new sounds and entirely new methods for making music required new tools. As Varèse remarked early on, 'Our musical alphabet must be enriched. We also need new instruments very badly ... which can lend themselves to every expression of thought and can keep up with thought' (1916, quoted in [207]).

This time in history holds many examples of the influence of electricity on instruments. Since electricity emerged musicians have readily and eagerly adapted it and other new technologies for creative ends [190, p. 253] The first example of an instrument that took advantage of newly-available electricity, the Denis d'or, appeared in 1753. Up until the early 20th century this application of electricity can be seen in various other electronic instruments, such as Grey's pioneering synthesizer (1876) and the vacuum-based Audion (1906). Though these are undoubtedly innovative adaptations of new technologies for use in music, these inventions were largely variations on and re-imaginings of an already-existing instrument, common in many homes: The piano.

Figure 2.1: Alexandra Stepanoff playing the theremin on NBC Radio, 1930 (Wikimedia Commons)

As electricity became more widely available innovation also increased in this space. In the early decades of the 20th century instruments began to emerge that were entirely novel — not only in their sound, but also in the way they were played. 1920 saw the debut of perhaps the most enduring electronic musical instrument in history: The Theremin. Developed by Russian inventor Léon Theremin, it consists of two antennae, each controlling an oscillator that influences either the frequency or amplitude of the output. By the player moving her hand closer to the antennae, the frequency of the associated oscillator, and therefore the pitch or amplitude of the output, increases.

The Theremin was used by professional musicians (such as the acclaimed performers Clara Rockmore and Alexandra Stepanoff, who is pictured in Figure 2.1) to perform classical compositions, well received by large audiences. Its use waned in the 1940s in favour of new instruments that were easier to play, but the Theremin has maintained a following of professional and hobbyist players up until present day.

The Theremin highlighted an important factor: Electricity saved labour in the home, but it also saved labour in music. Tiny, near-effortless movements could create vast variations in the pitch and amplitude of the resulting sound. Nearly a century later we are still debating the effect of this disconnect of musical effort and output in the context of DMIs (which I return to in Section 2.3).

Along with the development of entirely novel instruments, technology-led exploration began to break apart other established musical relationships. Pierre Schaeffer, led by emerging and available recording technology, became preoccupied with the musical 'object' that formed the basis of his *musique concrète*: The recorded sound fragment could, at last, be entirely independent from its acoustic source [112]. This separation of sound and its source is still a central question for DMIs, and I discuss it in detail in Section 2.3.

2.1.3 A survey of current DMI practice

This thesis refers often to the term 'DMI performance' to mean a particular kind of technology-led, experimental musical practice. This artistic community is prolific in its output considering its relatively small number of practitioners, but its experimental nature means that there is no easy definition of the term 'DMI performance'.

Part of the reason for this is that the practice is still maturing. Though electronic music experimentation in this form has been going on for a century, DMI practice as we know it now emerged in the 1990s with the wider availability of small computers powerful enough for musical applications. In the ensuing three decades there has been considerable creative, experimental, technical, and scholarly activity in this space, but there is not yet the level of critical discourse that defines other forms of artmaking. There are some works which serve to somewhat define DMI performance practice [141], but besides being extremely small in number these also focus primarily on the technological aspects of DMIs, leaving aside rigorous critical discussion of the artistic content of this art form, and most crucially, how DMI performance form impacts and is influenced by the artistic, social, and technical cultures in which it is made and consumed.

Perhaps this lack of canon is unsurprising, as DMI performance is, by its nature, radically experimental on all fronts and therefore resists formalised definition. Further, this community's lack of (or, perhaps more accurately, constantly expanding) boundaries mean that it is an elastic term that stretches to include a huge range of electronic musical practice that (usually) falls outside the bounds of established musical genres. This thesis does not presume that 'DMI performance' is a finite term, and this survey does not seek to exclude any practitioner or method of working. Rather, by citing a wide range of current working methods, sounds, approaches and practices that make up this community, I seek to demonstrate the diversity of its form, sound, practice, methods, and artistic goals.

Further, this section in no way serves as a complete or in-depth definition of DMI performance practice, either by historical or contemporary definitions — I have focused exclusively on current practitioners and recent work, and any exhaustive or historical survey is outside the scope of this thesis. This survey is defined by the time in which it was written, and it should be noted that any definition of DMI performance faces the challenge of the rapid evolution of tools, methods, aesthetics and practices that is so common within this community.

This rapid evolution means that many DMIs are designed and then disappear. Morreale et al. [144] point out that though a huge number of DMIs are produced every year just within the NIME community, few of these ever find a lasting life as a performance instrument and are not played again. The short life span of some DMIs, however, does not exlude them from usefulness or relevance — on the contrary, this radical, technology-led experimentation is perhaps the DMI community's most unifying characteristic.

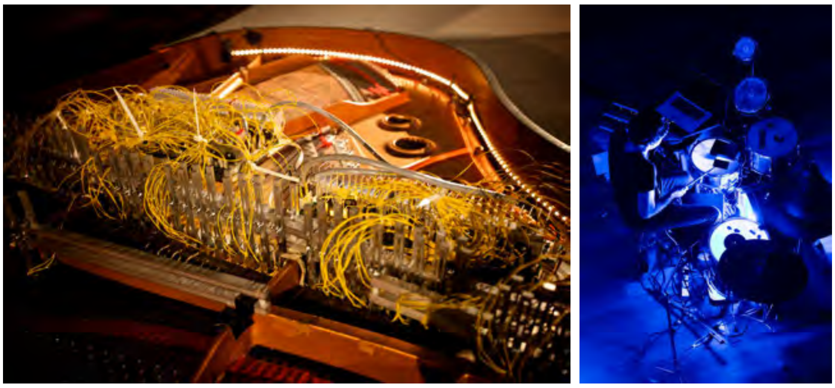
I also readily acknowledge that the performance of digitally-enabled musical instruments, methods and processes happens across all musical genres, from pop music to heavy metal. The term 'DMI performance', as it is used in this thesis, acknowledges these practices in other domains, but refers specifically to the experimental practices that are not primarily located within other musical disciplines. These categorisations are also not intended to truncate or limit the artistic practice of any of the above artists, and many can fit into several of the headings below and more besides.

Augmented instruments

The most obvious place to start with a survey of DMI practice is to examine instruments that extend, build upon, and augment existing traditional instruments in order to give them new capabilities.

McPherson's Magnetic Resonator Piano (Figure 2.2, left) is an example of this kind of DMI. Using 88 magnetic resonators and finger tracking, this system allows a piano to be played in its original form with the expected playing affordances, but presents a new dimension of timbral and playing possibilities [136]. These include, for example, the infinite sustain made possible by the magnetic resonators.

Figure 2.2: DMIs that are existing instruments that have been augmented. Left: The Magnetic Resonator Piano. Right: Christos Michalakos, augmented drums



Also notable is Christos Michalakos[1]'s augmented drum kit (Figure 2.2, right). This is an acoustic drum kit whose sound and visual presence is augmented in real time by a computer. While still being and functioning as a drum kit, this instrument offers an extended range of sound possibilities through live sound processing. This processing also drives the kit's embedded lights, lending a visual aspect that is absent from its traditional counterpart.

DMIs inspired by traditional instruments

As well as DMIs that are augmented versions of existing instruments, there are also DMIs that are novel but based on the sound and/or input methods of traditional instruments.

Dan Overholt's Overtone Violin [156] is one of these. This DMI's form factor references a violin, and it is played in a similar way. However, though it is made of wood, has strings and is bowed, this DMI is not a violin in the traditional sense: Not only does it have six strings, but it also offers the player a number of inputs and methods for controlling aspects of the resulting sound (see Figure 2.3, left).

The Push-Pull [96] is a more recent example of a DMI that references an existing instrument, in this case an accordion (see Figure 2.3). This instrument has a form factor and basic method of playing that references

[1] http://christosmichalakos.net/

Figure 2.3: DMIs based on existing instruments. Left: The Overtone Violin (from [156]). Right: Push-Pull, used with permission.

the accordion, but its sound and materials set it apart as a novel instrument[2]

DMI as networked object

DMI practitioners are also taking advantage of developments within the area of the Internet of Things to design instruments that challenge the formal constraints of audience/performer often found in DMI performance, and re-think the player-audience relationship. An example of this is DIADs (Distributed Interactive Audio Devices) by Bown et al. [33] (see Figure 2.4). These instruments are 3D printed egg-shaped objects, DIADs contain embedded loudspeakers (so the audience becomes the locus of the music),

[2]See a video of Push-Pull here: https://vimeo.com/110656141

Figure 2.4: DMIs as networked objects: DIADs. Bown et al. Used with permission.

Figure 2.5: Novel DMIs by composer-makers. Left: Laetitia Sonami performs on Spring Spyre (Photo credit: Brown University). Left: Myriam Bleau, Soft Revolvers (Photo: Devin McAdam, used with permission).

and sensors (so the audience interaction can influence the sound output). Performances generally use a group of DIADs, and the music is remotely controlled by a performer, and the audience interaction through the devices not only blurs the line between performer and spectator, but also encourages game-like exploration between the two.

Performance-focused DMIs

Some DMIs are musical instruments in the sense that they are objects used to perform music, but are not related to any traditional instrument's form or function. Rather, these instruments are created by composers for their own artistic ends, but also have a visual language of their own.

Laetetia Sonami's Spring Spyre is one of these (Figure 2.5, left). Composed of a large metal ring with springs stretched across it, aspects of the resulting sound are modified and modulated when the strings are rubbed, plucked, or stretched.

Soft Revolvers by Myriam Bleau is also a novel DMI that Bleau has designed to meet her artistic goals (see Figure 2.5, right). The 'revolvers' are acrylic discs with rounded bottoms that enable them to be spun like tops that make sound in response to their revolving behaviour.

Commercial DMIs

Though many DMIs do not find a life or playing community beyond the composers that make them [144], there are some DMIs that are not only innovative explorations of musical and interface possibilities but are also commercial products intended for use by DMI performers.

The MiMu gloves project[3] (Figure 2.6, left) has enjoyed a high profile thanks to early collaboration with musician Imogen Heap. The gloves, which allow the performer to record, play back, sculpt and interact with sound using on-board wireless communication hardware and sensors for gestural control, have been extensively used by Heap in a performance context and continue to be available for sale.

The Karlax[4] (Figure 2.6, right) is a purpose-built, high-quality instrument made for DMI performance. With particular attention paid to the form and design of the interface and attracting a handful of professional players, the Karlax has achieved a status usually not found by DMIs that are often built to meet a single composer's artistic goals.

Electronics-led DMI performance

The term 'DMI performance' also includes practitioners that make electronic music using a range of analogue and digital tools, some entirely self-built, and others hacked or re-purposed.

Lauren Hayes[5] (see Figure 2.7 is a musician and sound artist who performs on instruments that are analogue/digital hybrids. Her performances

[3]https://mimugloves.com/
[4]http://www.karlax.com/
[5]http://www.laurensarahhayes.com/

Figure 2.6: Commercial DMI applications. Left: Karlax. Right: Mimu gloves. Photos used with permission.

Figure 2.7: Electronics-led DMI performers. Left: Lauren Hayes; Right: xname. Photos used with permission.

combine elements of free improvisation, techno, and noise. Her interfaces, which incorporate elements ranging from analogue synthesisers to game controllers, are intentionally unpredictable.

xname[6] is a London-based performer who uses a combination of strobe lights, solar panels and analogue electronics to create ritualistic noise music. She uses hand-built circuits, using basic electronic components such as hex inverters, to generate analogue signal with rhythmic pulses of light that is sonified live on stage.

DMI as performance object

Some practitioners have been pushing the design and function of DMIs to the point where the instrument becomes both means of making sound as well as an aspect of the performance, an object with a performative presence of

[6] http://xname.cc/

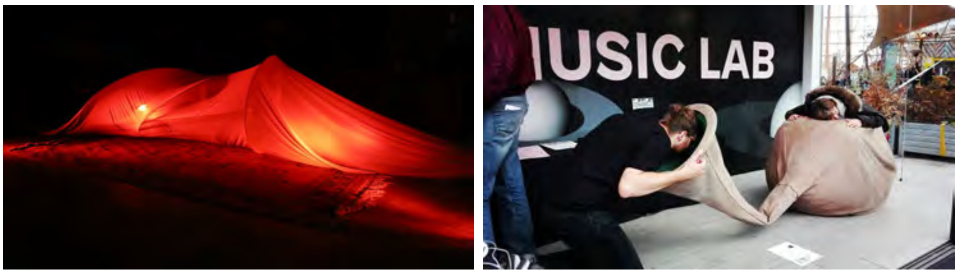


Figure 2.8: DMI as performance object. Left: Chrysalis by Marije Baalman. Right: Bellyhorn by Dianne Verdonk. Photos used with permission.

its own.

Marije Baalman's Chrysalis (Figure 2.8) is an example of this. In this performance the tent is a performance prop, visual communicator, and metaphor, as well as a DMI. The cocoon on stage contains the performer who is controlling the music from inside, her tiny physical movements sonified by a range of in-built sensors.

Dianne Verdonk's Bellyhorn (Figure 2.8, right) is another DMI that is part instrument, part sculpture. It is both performed by Verdonk and presented as an installation where viewers can try out the horn for themselves. When a participant sings into the 'mouth' of the instrument, their vocal input is processed and amplified through a speaker in the instrument's soft 'belly'. The haptic feedback of Bellyhorn's internal speaker gives it the added aspect of social sculpture, as audiences often lean on or against it to feel the vibrations as the performer's vocal output is amplified.

Software instruments and computer languages

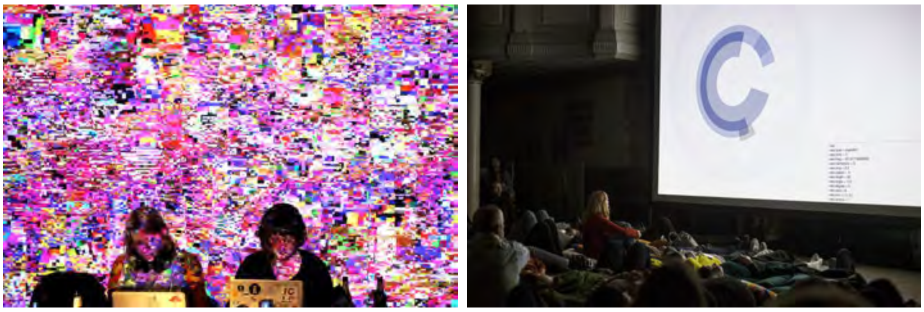
DMI performance also includes performances that are entirely software-based, expanding the notion of 'expressivity' through language to include expression through computer code.

Live coders use their laptops as sound-producing instruments and performance interfaces, usually projecting their code behind them as they work to provide a visual link between their actions behind the computer and the sonic outcomes. Joanne Armitage[7] and Shelly Knotts[8], associated with the Algorave[9] community, have recently collaborated to form Algobabez[10], a live coding female algopop duo (see Figure 2.9, left).

Thor Magnusson's Threnoscope [132] is an extension of the established live coding ethos. The Threnoscope, a microtonal live coding system, is not only a software-based DMI for producing and performing sound, but also produces visual output in the form of circular scores that visually evolve in real time in response to the sound produced (Figure 2.9, right).

[7] http://joannnne.github.io/
[8] https://algorave.com/shelly-knotts/
[9] https://algorave.com/
[10] https://twitter.com/algobbz

Figure 2.9: Software instruments. Left: Algobabez (Photo credit: Antonio Roberts, hellocatfood.com). Right: Threnoscope (photo used with permission).



Body-focused approaches

There are some DMI performers who treat the sensor as an instrument, which extends — and is inseparable from – the body that plays it.

Atau Tanaka has pioneered performance in this area through his investigation of sensor performance. BioMuse is an instrument based on physiological sensing technology, and a combination of 'cerebral activity (electroencephalogram, EEG), cardioid activity (electrocardiogram, ECG), and muscular activity (electromyogram, EMG) are sensed and digitized, and become human interface data for the articulation of computer processes and media such as digital audio, video, and computer graphics' [191].

Onyx Ashanti[11] (Figure 2.10, right) creates self-designed, 3D printed instruments that expand and extend the musical functionality of his body in order to explore his self-styled genre of 'beatjazz' — beats being electronically produced rhythm, and jazz referring to an improvisational exploration of musical possibilities. Capitalising on emerging technological trends such as 3D printing, crowd funding and open-source technology for creative purposes, Ashanti positions his technologically-augmented body as both the place and means of musical creation.

Synthesis as instrument

Synthesis and related computational techniques are core ways to create sound with a DMI, and have been explored and applied to DMI performance in a number of ways. One of these is Victor Zappi's Hyper Drum-

[11] http://onyx-ashanti.com/

Head (Figure 2.11), a touch interface where the performer, using the 42" transparent projection screen, can draw the outlines of 'rooms' and make sound by synthesizing the impulse response of that room by tapping on it [213]. In drawing multiple 'rooms' in real time, the performer can create a percussion surface that uses this real-time synthesis in a variety of ways, the interface evolving over the course of the performance. This real-time synthesis is made possible by innovative application of GPU hardware and programming.

Other experiments

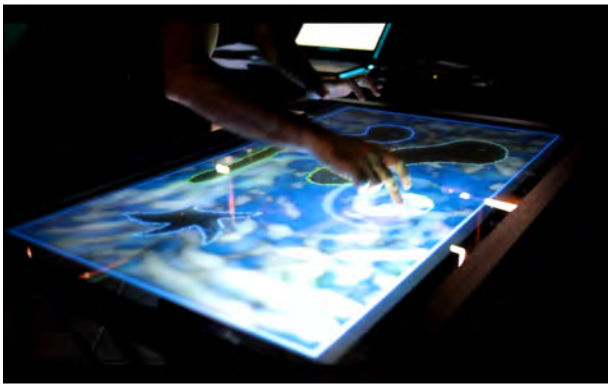
The limits of DMI experimentation, as well as the artistic content that comes from how this music is produced, are continually being pushed by a subset of practitioners who are rethinking what it means to make music and what an instrument can be. These explorations are often highly metaphorical, and meditate on the influence of technology and its meaning and purpose.

John Bowers and Benjamin Freeth explore this in their performance of Bio-Vortex (Figure 2.12, left), which explores 'wet machines' as an interface for sound (in this case that wet machine is an assemblage of electronics, software, and bioluminescent algae). The algae's natural metabolism is light-producing, which is in turn detected by light-sensitive electronics, blurring the line between natural/mechanical, human/animal, analogue/digital. By feeding a living thing, it in turn gives 'life' to a digital system.

Martin Howse creates performances such as The Dark Interpreter (see Figure 2.12, right) in which he uses dirt, earth and other naturally-occurring materials as an interface for sound. This method of music making is expres-

Figure 2.10: DMIs that are body-focused. Left: BioMuse (Atau Tanaka). Right: Onyx Ashanti during his TED talk [8]

Figure 2.11: Synthesis DMI: Hyper DrumHead performed by Victor Zappi.



sive of Howse's artistic goals, which are to investigate the link between the earth, software, ritual, and the human psyche.

2.1.4 What 'DMI' means in this thesis

In this section I have provided a tracing of DMI's historical roots, and a survey of contemporary performance practice. This survey demonstrates the breadth of diversity in this artistic community, as well as the range of methods, tools and approaches that the term 'DMI' can include.

I will discuss the specific questions and considerations emerging out of DMI research later in this chapter, but I pause here to articulate what 'DMI' refers to in this thesis. With such a huge range of methods, processes, forms, and practices, one PhD is insufficient to equally consider them all, and a limit of scope is necessary.

The DMIs that are the focus of inquiry in the three studies this thesis is

Figure 2.12: Experimental DMIs. Left: Bio-Vortex by Bowers and Freeth [171]. Right: Martin Howse, The Dark Interpreter [170]. Both images used with permission of the artists.

limited to DMIs with *instrumental interaction*, as defined by Cadoz [39]. As demonstrated above, a huge range of things can fall under the term 'DMI', and Caodoz proposes that the term 'instrument' has been overstretched when applied to DMIs. He defines the term 'instrumental interaction' as referring to a subset of DMIs that require mechanical processes through which there is energetic exchange.

DMIs with 'instrumental interaction' are the primary consideration in this research. This is not to suggest that the outcomes are not useful to the design and study of other types of DMIs, but instead indicates that the studies in this thesis are focused on audience perception of instrumental interaction, and it is these DMIs on which the findings are based, and the primary area of application.

2.2 How audiences experience music

A primary goal of the studies in this thesis is to understand how audiences perceive, judge and experience musical experiences. This section surveys some existing theories, from domains including music and HCI theory, as well as further afield.

2.2.1 *Werktreue* and the score

Philosopher Lydia Goehr lends insight to this question through her theory of *Werktreue* [83]. According to Goehr, *Werktreue*, or 'the work-concept', emerged at the end of the 1700s, and was originally a way of standardising and governing the relationship between the composition and the performance.

Werktreue, Goehr suggests, emerged as a way to address the varying interpretations of musical graphic notation, in order to ensure that music would be able to be appreciated in its 'true' form in the future. Entwined with this notion was the 'duty' that the performer had, to produce a 'true' representation of the work, as well as the composer's intention[83, p231].

In a time before recorded music, the utility of such an agreed-upon system is obvious: For example, a standardised way of playing and understanding music meant that one composition could be sent across great geographical distance and reliably reproduced elsewhere. However, *Werktreue* has since achieved a status so pervasive and established that it can be described as hegemonic.

Notable within this context is the work of John Cage, which opened up musical possibilities by challenging widely-held notions of music. Though Cage recognised the potential of novel instruments for musical experimentation and is noted for inventing and re-purposing musical instruments (most famously the prepared piano), his motivations were not akin to the heady techno-lust of early modernist practitioners. Instead, his motivations were far more pragmatic. As he remarked: 'Technology essentially is a way of getting more done with less effort.' [121] He did, however, bemoan the lack of innovation being undertaken by composers, despite a wealth of new tools:

> *When Theremin provided an instrument with genuinely new possiblities, Thereministes did their utmost to make the instrument sound like some old instrument, giving it a sickeningly sweet vibrato, and performing upon it, with difficulty, masterpieces from the past ... We are shielded from new sound experiences.*[41, p. 4]

Cage's scores often incorporated the element of chance, and he embraced the indeterminate outcome. His piece 4'33' (1952), for example, consists only of silence laced with the incidental stirrings of the concert audience. Cage also produced a number of visual scores, which do not conform to rules of musical notation (and sometimes do not use any musical notation at all), and require the performer to apply their own interpretation (see Figure 2.13). Gurevich identifies this indeterminacy presenting new musical

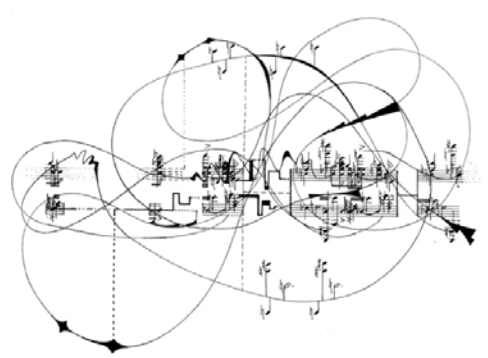


Figure 2.13: An example of a Cage graphical score

possibilities via a method which 'diminishes the composer's own role in the sound-making process, providing performers ... with a limitless possibility space within which to make creative decisions.' [85]

It is important to stress that though Cage was actively challenging accepted notions of music, he was still in some aspects conforming to *Werktreue*. Goehr writes:

> *Despite the apparent absence of pre-determination in the composition, despite the experimental nature of performances where the emphasis has deliberately been placed on natural sounds rather than on intentionally produced sounds, Cage has not obviously succeeded ... in undermining the force of the work-concept within the musical institution. First, he has maintained control (however minimal) over the music. It is because of his specifications that people gather together, usually in a concert hall, to listen to the sounds of the hall for the allotted time period. In ironic gesture, it is Cage who specifies that a pianist should sit at a piano to go through the motions of performance. The performer is applauded and the composer granted recognition for the 'work'. Whatever changes have come about in our material understanding of musical sound, the formal constraints of the work-concept have ironically been maintained.* [83, p. 264]

In this way, Goehr makes a distinction between *Werktreue* and what may be called **musical vernacular**. Though vernacular and accepted methods of musical communication are aspects of *Werktreue*, the concept denotes a much broader and deeper hegemony that incorporates a wide array of aspects, among them audience behaviour, the role of a 'composer' and the privileging of his or her ideas, and so on.

Nevertheless, Goehr states that Cage left the 'formal constraints' of *Werktreue* intact, while specifically undermining the work-concept of the score as the singular ideal [83, p 164]. Gurevich notes that although he embraced the indeterminate outcome, Cage still reacted with anger to 'perceived abuses of his openness' [85], suggesting that instead of creating a space of infinite musical possibility he instead still had very specific ideas of what a given 'work' should be.

This challenging of the score and musical vernacular while leaving formal

Figure 2.14: Score for Age by Bill Drummond (used with permission)

constraints intact is evident up to the present day. Bill Drummond, famous (perhaps ironically) for his production of pop music as half of the KLF, produces scores in a consistent graphical style that ask the performer to listen, remember, think and feel as a musical act (Figure 2.14). Drummond, despite disposing of musical vernacular entirely, still retains the formal constraints in some aspects, such as by naming works as 'scores', by stating that they are to be performed, even sometimes providing details of location, method, and duration of the performance.

Differentiating between musical vernacular and formal constraints, however, can offer some insight into the progression of electronic and digital music into the cultural mainstream. Though transparency of DMIs is a much-discussed topic in NIME research (and I return to this in considerable depth in Section 2.3), the musical mainstream has readily and enthusiastically adopted electronic and digital instruments and digital processes into popular consciousness. The Theremin appeared in popular culture as early as 1951 on the soundtrack to *The Day the Earth Stood Still*, but by 1968 Brian Wilson was playing an adapted Theremin (called a Tannerin) in the Beach Boys' number one *Good Vibrations*. At the same time synthesizers were entering mainstream consciousness, with Wendy Carlos achieving considerable acclaim for her 1968 album of Moog interpretations of Bach pieces, *Switched On Bach*, as well as her synthesizer-based soundtrack for Kubrick's 1971 hit *A Clockwork Orange*. The synthesizer was even introduced to children via Bruce Haack, who demonstrated his own invention in a suitcase on *Mister Rogers Neighbourhood*. As the 70s and 80s progressed the synthesizer became a mainstay of pop music, and digital sounds, beats, processing and recording techniques, as well as the MIDI (Musical Instrument Digital Interface) protocol, have been wholly adopted and used by the music industry at large. In short: The mainstream has not rejected electronic or digital music; if anything, the mainstream has demonstrated a voracious and continuous hunger for all things digital.

However, DMIs stand apart, having never gained the commercial success of their mainstream counterparts. Commercial success and popular taste are anything but reliable metrics of artistic worth, but the differences between the two ways of making digital music are notable: The electronic and digital instruments that are novel and have been widely successful — such as those employed by German innovators Kraftwerk — retain musical vernacular. (This idea is explored in more depth in Chapter 5, where I question the role of familiarity in DMI performance.)

I will refer to this notion of 'musical vernacular', as described here in the context of *Werktreue*, often in this thesis. I use this term as a shorthand for the aspects of musical convention (such as pitch, rhythm, melody, playing style), and excluding what Goehr terms 'formal constraints' [83, p. 264], (the concert venue, seating, performance context, duration).

2.2.2 Models for understanding audience perception of music

In order to understand how audiences understand and make sense of music, a number of models have emerged from the fields of music perception, DMI research, and HCI. Here I survey some of these, and describe their useful and less useful aspects when applied to DMI performance.

The ITPRA theory of expectation

Huron [101] states that the performing arts have been a primary area of study by those who aim to understand the dynamics of emotion. His ITPRA model describes five 'expectation-related emotion response systems' [101, p 15] known as ITPRA, an acronym for the responses this model describes. Two, *Imagination* and *Tension*, precede the event, and the following three, *Prediction*, *Reaction* and *Appraisal*, follow it:

> **Imagination:** A continuum of expectation generated in response to our environment.
> **Tension:** Optimum arousal and attention in preparation for anticipated events
> **Prediction:** Negative/positive reinforcement to encourage the formation of accurate expectations
> **Reaction:** Neurologically fast responses that assume a worst-case assessment of the outcome
> **Appraisal:** Neurologically complex assessment of the final outcome that results in negative/positive reinforcement

This model is useful as it breaks up the spectator experience along a temporal line, and identifies that different processes execute as a musical experience unfolds. However, the emphasis on expectation also opens up questions about how audiences react when there is no opportunity to form expectations of music - such as in DMI performance, which rejects musical vernacular and therefore provides no opportunity for expectations to form.

Huron engages with this, and asserts that the work of modernist composers such as Wagner and Stravinsky 'only make sense when viewed as serving the goal of psychological disruption through thwarted expectation.' [101, p. 350] In other words, their rejection of musical vernacular was still reliant on the expectations of musical vernacular that they knew their listeners would have, and were operating using the psychological mechanisms of expectation in place, and that 'Schoenberg wielded the same psychological tools of expectation that Mozart did, even if he sculpted very different works.' [101, p. 353]

While intriguing, this does assume two things: Firstly, that all composers have a clear intent to convey a specific, knowable emotion, and secondly, that the psychology of expectation is somehow inherent in humans and not a learned behaviour, and will therefore always be there.

The Implication-Realisation Model

Narmour's Implication-Relization (or I-R) model describes the role of expectation in melody. This model proposes that there are two independent but interacting systems of musical expectation [147]:

> **The 'top-down' system** is where a listener perceives music and compares it against their own personal frame of reference. In this system, both prior learning before listening, and learning while listening, influence expectations.
> **The 'bottom-up' system** is an automatic and innate system which perceives music as if it is hearing it for the first time.

Further, in his major work on the I-R model [148], he cites two hypotheses that are central to this notion. First, that repetition implies further repetition, and second, that contrast implies further contrast. These two hypotheses, according to Narmour, are central to understanding the typical methods of expressing musical structure.

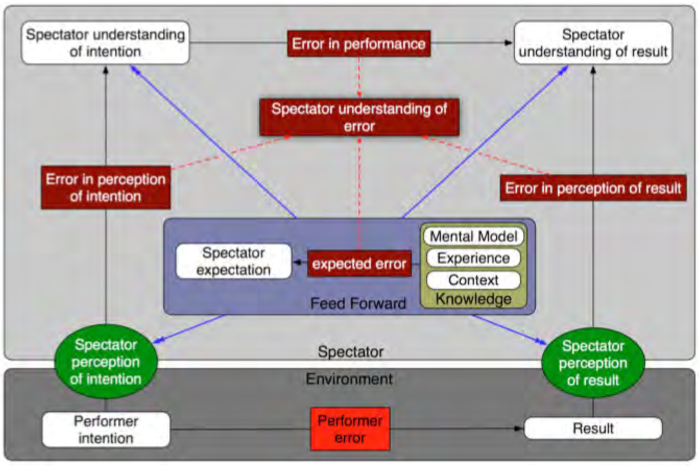
It is important to note that Narmour's model sits definitively in the realm of melody analysis, and therefore can't reliably be generalised to understand audience perception in a live context. It is, however, useful to separate the individual from larger, intrinsic forces that may be working upon their perception.

Understanding of audience perception of skill and error from the NIME community

There has been a considerable amount of investigation done on the spectator experience in NIME by Fyans and Gurevich and collaborators. In this section I summarise these findings on audience perception of error and skill, and the models developed to describe each of these phenomena.

The earliest of these works proposes that, despite NIME's emphasis on transparency (a notion that is more deeply examined in Section 2.3), the mental model of the spectator is a better tool for examining how the spectator understands and assesses a musical performance [74] (Figure 2.15).

Figure 2.15: Fyans et al.'s model of error [74]



The authors state that a spectator's understanding of 'success' or 'error' (which, notably, they position as binary opposites) is dependent on this mental model, and a spectator perceives 'success' through the closeness of their understanding of the performer's intent to the understanding of the result.

In a related publication [73], the authors suggest that there are five questions (adapted from Bellotti et al.'s five questions for designers of interactive sensing systems [20]) that a DMI designer must consider for spectators to perceive skill in performance:

1. **Address:** How does the spectator know that the performer is directing communication to the system?

2. **Attention:** How does the spectator know that the system is responding to the performer?

3. **Action:** How does the spectator think the user controls the system?

4. **Alignment:** How does the spectator know that the system is doing the right thing?

5. **Accident:** How does the spectator know when the performer (or the system) has made a mistake?

The authors focus on question 5, Accident, and their model (Figure 2.15) describes the kinds of error that may arise, and their magnitude (based on the length of the line on which they sit). They propose that though this model was developed in the context of DMI performance, it can be generalised to any performance, such as traditional music or sport.

Notably, the authors make a distinction between *actual* error and *perceived* error, suggesting that error may exist whether the audience notices or not.

Most importantly, this model not only includes the idea of error 'magnitude', suggesting that some errors are more detrimental than others. But, a discussion of what this detriment is, exactly, is absent, and the mutual exclusivity of this success/error relationship lends no clues. This model assumes that everyone wants success and no one wants errors, but does not query if this is actually the case.

Further, without musical vernacular and with no cohesive performance conventions in the DMI performance tradition, the subtleties of intent, success, and error become very murky, and the influence of HCI methodologies in NIME that override musical considerations (I discuss this influence further in 2.3) become starkly clear. This may be a model that can describe how spectators perceive a wrong button press, or a misplaced gesture, but it does not engage with what these errors mean for that spectator's aesthetic and musical experience, or if it even matters.

These publications do, however, highlight the importance of the spectator's mental model when considering the spectator experience of music. This is consistent with literature in the music and music cognition domains, which also cite mental models as key to understanding music [50, 159]. In a later and related publication, Fyans et al. conducted a test of various spectators, matching their understanding of aspects of the DMI domain (music, interaction, electronics) and their accuracy in spotting 'errors' [75]. They concluded that when a spectator lacks an accurate mental model of a performance, they rely instead on facial and gestural information instead to judge errors, and recommend making errors more obvious to improve spectator undersatnding.

Despite this being a NIME-specific investigation, conspicuous by its absence in these two publications is engagement with the factor most confounding for audiences of DMI performance: The musical content. It is precisely DMI's lack of musical vernacular that prevents the audience from even *forming* stable mental model, let alone have one to rely upon for judging musical events as they unfold in real time.

Most intriguingly, Fyans et al. posit in [75] that their test DMI that was entirely novel in its interaction, the Tilt-Synth, had very few errors associated with it, and that some participants suggested that it was, in fact, **impossible for the performer to make any error at all**. This suggests that the spectators are indicating an absence of mental model for this kind of DMI performance: There are no boundaries beyond which to stray, no vernacular to defy. I will return to this in a discussion of error later in this chapter, in Section 2.5.

Fyans and Gurevich in 2011 published a study of how audiences perceive *skill* in DMI performance [72]. This study was motivated by the community desire for DMIs that support virtuosity. They state that within NIME, despite this hunger to be virtuosic, there is little discourse around what 'skill' means in a DMI context.

They conclude that there are factors that contribute to an increased perception of skill in DMI performance that are communicated by the performer, such as confidence and embodiment, and they reduce skill to a combination of two things: Control and effort. This is a definition I return to throughout this thesis. They also suggest social structures that might reinforce skill, such as 'a community of practice', suggesting that in order for the performer's skill to be meaningful the spectator must have knowledge of this specific domain — a conclusion that again questions the usefulness of the absence of musical vernacular in this performance domain, and seems to suggest imposing limits.

But, this notion of skill as a combination of control and effort that must be perceivable by the audience is useful, and aligns with the 'effort heuristic' proposed by Kruger et al. [122], which they describe as commonly-used, 'general judgmental tendency' that equates greater effort as perceived by the viewer with greater skill on the part of the artist, and greater subjective and objective quality.

Expressivity and the 'emotional pipeline'

'Expressivity' is a much-used word in DMI and HCI literature, but no single definition exists. This is not a problem unique to DMIs - indeed, attempts to understand performer/audience communication far predate the digital age [176]. Fels et al [68] use the term to mean the communication of meaning or feeling. Expressivity is generally agreed to include an aspect of *communication* and an aspect of *emotion*.

Poepel [164] describes this communication of meaning and emotion as a process of encoding (by the performer) and decoding (by the audience). He suggests that the methods of performer encoding are tempo, sound level, timing, intonation, timbre, vibrato, tone attacks, tone delays and pauses. Juslin [111] acknowledges that the listener may well influence a performer's expression through factors that may be piece-, instrument-, performer-, listener-, and/or context-related. These communicative models are grouped by Murray-Browne [146] into the concept of an 'emotional pipeline', a medium where music is used as a signal to transmit emotional messages.

This model of musical expressivity, however, is limited to scored music. Dobrian and Koppelman conclude, in their critical review of expressivity in DMIs [62], that if DMIs are to achieve this kind of expressivity then a community of players, audiences and repertoire is necessary to establish a common language.

It is entirely possible that this kind of communication simply isn't realistic without genre-specific knowledge on both the side of the performer and the audience, but this is precisely the kind of hegemonic structure that Cage and other musical innovators have been challenging. If DMI music is defined by technology-led exploration that rejects all boundaries and vernacular, it is illogical to presume that imposing vernacular on this performance tradition will resolve this long-standing problem, as a lack of boundaries and vernacular are the defining characteristics of DMI performance.

In 1996 Gabrielsson and Juslin argued that, rather than establish the 'lawful' relations between the score and the performance, one should instead study 'the relations between the performer's intentions, the variables in the sounding music, and the listener's experience' [76]. In 2007 Gurevich and Treviño [89] proposed a similar approach specific to the NIME community. Hesitating to limit DMI performance to being a 'representational art', they stated that it is'hegemonic' to demand that DMI music fit within the bounds of the classical Western tradition in order to be considered successful. Instead, they suggest that an *ecological* approach to DMI music, which privileges the relationships between all parties involved (performers, audience, composers, and so on) and considering fully the specific context of these relationships. (I return to a discussion of this ecological approach in Section 2.3.)

Dervin's sense making methodology

Brenda Dervin's model of 'sense making' [55] is a further model that may be useful in understanding audience perception in the context of DMI performance.

The above models presume that sense is made through mental models, that are formed before and during a performance, which create the basis of expectation. This expectation, in turn, allows audiences to continuously judge and evaluate musical events against their expectations.

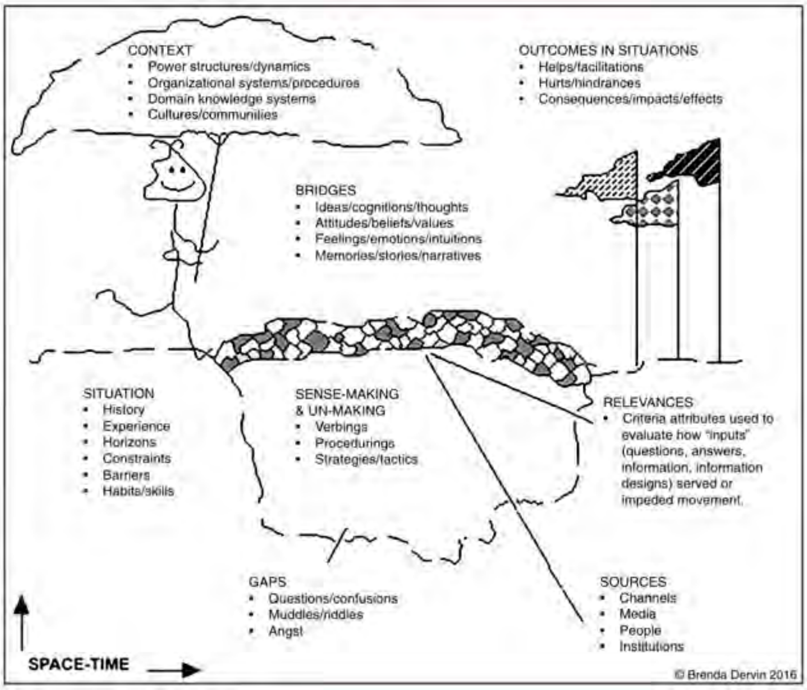
As explored earlier in this section, however, the fact that DMI performance favours technology-led radical exploration over the reinforcement of musical vernacular presents a major confounding factor for audiences. Without any shared knowledge, without a community of practice, and with experimentation with vernacular a core value of the DMI performance genre, taking these theories from the world of scored, vernacular music to this radically experimental one may not be possible.

Sense making may lend insight into this confounding situation. Coming from the field of knowledge management - a field primarily concerned with library information science - Brenda Dervin's sense making methodology (SMM) has applications here. Dervin considers knowledge management to be 'a symptom of, and a proposed solution for, human confrontation with issues of chaos versus order and centrality versus diversity.' [54] SMM, instead of assuming or requiring a static system of established knowledge between two parties, instead acknowledges and embraces the inherent chaos of a world in flux. Says Dervin:

> *While once we thought we could bask in the certainty of answers and solutions, now need to learn to appreciate the courage and creativity it takes to step into the unknown only partially instructed by information/knowledge. In this view, every next moment is unknown; and the step into it can never be more than partially informed.* [54]

Central to sense making is the idea that we can think about a complex plurality of ideas without resorting to hegemony, completeness, or imposing

Figure 2.16: Dervin's sense making theory, updated for 2016 [208]



a static structure. Although it may appear to be reaching to apply knowledge management theory to audience perception of DMI music performance, Dervin's sense making methodology (SMM) has been in development for over 40 years, and has been applied to audience studies of all descriptions, including the arts [208], HCI [173], and computer science [118].

Though an in-depth exploration of SSM is beyond the scope of this chapter, it is useful to examine the differences between this theory and the mental model and expectation theories described above. SSM (illustrated in Figure 2.16), posits that sense making is the process through which a person traverses gaps in knowledge or understanding. At its core, this model begins with a context/gap/outcome triangle, suggesting that a person begins in a situation, with their own context, and builds a bridge to traverse the gap to achieve an outcome.

The bridge is the central structure in this model. In other theories of music perception, this bridge is dependent on the spectator's knowledge of musical vernacular, and/or the shared musical language between performer and spectator, and suggests that all bridges would be identical, or that there is, at the risk of over-extending this metaphor, one *best* bridge. However, SSM adds nuance to this model: The bridge is composed of the spectator's own ideas, thoughts, feelings, memories and so on that may arise during the experience, which form the basis of their understanding. This viewpoint aligns with Tanaka's idea of experience as personalised: 'Experience is defined to be personal and self-referential, and implies that an individual can be proactive in shaping its own destiny.' [189, p. 282]

In this way, expressivity becomes less of a transactional 'emotional pipeline', and a more individualised journey. SSM adds further and useful nuance to the existing models of musical perception and expressivity, precisely because it does not assume that this communication is only possible in one way, and instead assumes a plurality of methods. This is therefore a model that is potentially very useful when considering the audience experience of DMI performance, given that the shared language of musical vernacular is missing in this context.

2.3 Musical instruments in the digital age

By the 1990s, the availability of powerful, affordable, and small computers overtook analogue technology, and as a result DMIs have become commonplace. [47] A DMI is defined as 'an instrument that contains a control surface ... and a sound generation unit' [198], or sound-generating computer. Mapping between the interface and computer acts as the glue that connects these two independent elements.

A 'musical instrument' has a very specific cultural connotation with millennia of history. As Tanaka specifies:

> *The term musical instrument has a clear connotation across many cultures. An instrument is imagined to be a known physical apparatus that allows human performers to express themselves artistically through sound. Musical instruments in the traditional sense are assumed to be acoustic, constructed of wood, metal, and other materials, having resonant qualities. Sound is articulated when the user intervenes and excites vibrational modes. Music is made through skilful manipulation of the instrument, resulting in melody, harmony, and rich sonic timbre.* [189, p. 268]

Tanaka reminds us that 'music has always pushed the envelope of what defines interaction' [189] and this section does not suggest that acoustic instruments as ossified artefacts that have been the same since time immemorial. That said, the shift presented by DMIs is profound. Though the disconnect of the instrument from the sound it makes has distinct advantage of a huge range of sonic possibilities for the DMI performer, it is precisely this disconnect that is the source of confounding factors for the audience.

Departing from Tanaka's definition above, several differences can be isolated that are areas of debate in DMI research, and areas which present confounding factors for audiences — though I discuss them in the context of DMI research, they relate directly to the presence of a computer and are, by extension, also HCI issues. In this section, I discuss each of the following:

- **Transparency.** Unlike acoustic instruments, DMIs are not necessarily resonant, and their sound is not necessarily dependent on their materiality and/or form factor. As a result, the way a DMI is played may not reveal to the audience anything about the relation of the performer gestures and manipulations (inputs) are related to the resulting sound (outputs).

  This issue has been termed by Fels et al. as *transparency* [68], and is a source of vigorous discussion within the DMI community.

- **Performer agency.** Related to transparency, a DMI is not always a human expressing themselves through sound. Because of the computer that by definition exists between the input and output, audiences are often confounded by which element - human or machine - is making the creative decisions.

- **Instrument familiarity.** The term 'digital musical instrument' does not have a clear connotation across many cultures, and is not a known physical apparatus. Because of the rise of the composer-maker and DMI performance's radical, technology-led experimentation, a DMI often does not allow audiences to form any expectations based on the nature of the instrument.

Finally, I discuss the influence of HCI research and thinking on how these questions have been approached by the DMI research community, which is largely located within the HCI research community. I discuss why HCI approaches, though they have lent much insight into the performer-instrument relationship, have not resulted into much insight at all into the audience experience.

2.3.1 Transparency: Perceiving cause and effect

For a DMI, only computational mappings link the interface and the output, and mappings that can be extremely complex, if not invisible. It is therefore difficult, if not impossible, for an audiences to understand how the performer is making sound, and therefore impossible for them to infer the performer's intent (the importance of forming mental models and understanding intent was discussed in Section 2.2).

The DMI research community has long been aware of this fundamental problem of the relationship of the inputs to the outputs outputs being invisible to audiences, and discussion has largely been centred around the notion of **transparency**. In 2002, Fels et al. defined the notion of 'transparency' in this way:

> *[T]ransparency provides an indication of the psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping. The more transparent the mapping is, the more expressive the device can be.* [68]

At its root, transparency describes the quality of the understanding of the relationship between input and output. Fels et al. suggest that transparency for the performer - which Moore terms *intimacy* [143], a term that Fels aligns with as well [67] - means that the output perfectly matches their expectation and control. For the audience, this means that the spectator understands which controls produce which sound.

According to Fels et al, increasing transparency, through clarifying the relationship of input to output, will increase expressivity (the authors do not elaborate on the meaning of this term, though I discuss it in Section 2.2). This perceived expressive deficit has been an influential idea, and the NIME community has since then produced a number of attempts to address this perceived lack of transparency, such as popularising instrument demos before performances to familiarise the audience with an instrument, making the gestural inputs visible [160], and visualising signal flows within an instrument [24].

Though useful in highlighting the confounding divide inherent in DMIs between the gesture and the output, the transparency model does not provide a blueprint for understanding or the audience experience of DMI performance. Jordà notes that complex mapping is not only the domain of DMIs: Woodwind instruments, for example, are extremely complex, with aspects controlling more than one parameter of the resulting sound (how hard the player blows not only controls the amplitude, but also the pitch and timbre of the output), but the discourse around how to make acoustic instruments more understandable are conspicuous by their absence [110].

The transparency discussion, at its root, reflects a view of musical performance that places the performer interaction at the apex of a number of factors that affect audience understanding and, by extension, enjoyment. However, live music performance is a multimodal experience, and sensory and extra-musical factors profoundly affect audience perception. It is reasonable to wonder, then, if instrument transparency has as profound an impact on audiences as has been assumed — larger than any of the many factors at play.

One aspect of this multi-modality that is directly linked to transparency is *gesture*. DMIs, unlike acoustic instruments, have no physical-world connection to the kind of sound produced (a phenomenon known as **control dislocation** [141]). As a result, huge sound outcomes can be the result of minuscule gestures that may not be perceptible by the audience, and may require almost no performer effort. This is a profound difference between DMIs and acoustic instruments: As Schloss points out, for thousands of years the way an instrument is played and the way it sounds have been intrinsically connected, and only in the last 30 years has this causal relationship broken down [175].

It is unsurprising, therefore, that the NIME community is intensely interested in gesture. In 2014 Jensenius found that NIME uses the word 'gesture' in an average of 62% of publications per year, far more than other related fields (SMC: 34%; ICMC: 17%) [105]. Despite this intense scrutiny, 'gesture' does not have one singular definition [40], though a workable notion definition comes from Miranda and Wanderley's 2006 concept of *instrumental gesture*:

> *[I]t is applied to a concrete (material) object with which there is physical interaction, and specific (physical) phenomena are produced during a physical interaction whose forms and dynamics can be mastered by the subject.* [141]

Gesture is often cited as a component of this larger notion of transparency, because gesture can provide the audience with knowledge of the way in which the inputs a performer uses relate to the sound outputs of a DMI. Fels et al. [68] suggest employing metaphor as a way of helping the audience build that mental model of the DMI's internal mapping, and gesture can be used as a method of demonstrated, metaphorical communication.

Despite this consensus that gesture is important, there is not much indication of how an effective expressive gesture might look, or how we might craft interfaces to support effective gestures. Schloss [175] further states that a 'visual component is essential to the audience' and that 'effort is important', but goes no further. Visi et al. [202] suggest that there exists a gestural vernacular of sorts, in that gesture is a shared language between audience and performer. They describe this as a 'layered process of signification — situated in a cultural ecology and shaped by shared knowledge' [202], but do not elaborate on what that language might be.

A study by Tsay [194] offers insight into gestural vernacular, from the classical music context. A group of musical experts and musical amateurs were asked to judge entrants in a classical music contest. When amateurs and experts listened only to the audio recording of the contest, both groups were not very successful at (12% for amateurs and 20% for experts). However, they watched silent video footage of the contest, both groups fared much better (about 46% success rate across both groups). This study suggests that the visual component an essential part of the musical experience, and that musical quality in this domain is communicated — and perceivable — through the visual components of performance.

In the DMI domain, a study by Bayatas et al. [17] investigated spectators' experience of tension in live-sequenced electronic music. This study found that the perception of emotional intensity, indicated with a linear potentiometer, is consistent across hearing and sight. Bayatas's results indicate that emotional intensity is similarly communicable through the audio and visual aspects of performance (though it should be noted that this 'performance' consisted of video of a person standing at a console in a room, and not footage of an actual performance).

What these studies suggest is that as well as a musical vernacular and audience expectations of what they hear, there may also exist vernacular and expectation in other modes of audience perception. The focused discussion on gesture, much like the discourse around transparency, does not make much allowance for other aspects of performance that may contribute to understanding. Curiously missing from this wider discussion entirely is engagement with what is perhaps the most confounding factor for audiences: **The music being played.** With the absence of musical vernacular, it is impossible to say whether more transparent interfaces or more demonstrative gestures would create more enjoyable performances. (This disentanglement of the influence of familiarity of both the instrument and the musical style is the subject of the first component of this research, the study described in Chapter 5.)

2.3.2 Instrument familiarity: The rise of the composer-maker

Where DMI performance lacks established musical vernacular that might enable spectators to form expectations, this continuous innovation in DMI design may be similarly confounding. Morreale et al. [144] surveyed the creators of instruments presented at NIME between 2010-2014. They found that the the greatest portion of instruments had been performed in public 0 or 1 times (23.5%), and that over half of instruments have been played by fewer than 3 musicians.

Though this inquiry is limited to a sample of DMIs featured at one conference over a five-year period, it does highlight an important confounding factor for audiences: DMIs are constantly being developed, and because of the absence of limits on their physical design or the sounds they can make, the audience can form no expectation based on what the DMI looks like or the way in which it's played.

In some ways, this is a symptom of this domain's emphasis on freedom: The composer-maker has long been a mainstay of DMI performance practice, innovating and experimenting through the design and creation of instruments for specific artistic and musical goals.

Fiebrink et al. [69] conducted a series of workshops for composers who worked with Wekinator[12], interactive software for to use machine learning for creating interfaces and compositions. The authors collated their observations on what composers valued, which they gathered through feature requests, complaints, discussions, and questionnaires.

This list of values is entirely performer-oriented. This is perhaps understandable in evaluations of software for music creation, but 'audience' and 'spectator' do not feature in any of the considerations. This is not to suggest that these performers are not at all concerned with audiences, but rather suggests that the the goals of DMI experimentation are device- and performer-focused, which may in some ways explain why the audience remains under-considered in the literature.

Much like the dismantling of musical vernacular, this continuous innovation has also disrupted the vernacular surrounding what an instrument communicates as an object. Dobrian and Koppelman [62] and Fyans et al. [72] have indicated that communities of practice are essential to expression with DMIs. However, the relentless pursuit of novelty through technological innovation, a modernist value that persists to this day, prevents the development of any cohesive understanding of what expectations may be formed around a given DMI based on its physicality.

2.3.3 Performer agency: Where the human stops and the machine begins

Though the relationship of inputs to outputs is certainly a confounding factor, more confounding still is the computer that mediates the relationship of those inputs and outputs.

At the centre of this factor are the notions of *liveness* and *agency* within DMI performance. I will deal with these each in turn.

Liveness is a term that, within DMI literature, subtly wields a double meaning. There is, first, 'liveness' in the sense of in-person performance, which Philip Auslander defines as 'the kind of performance in which the performers and the audience are both physically and temporally co-present to one another' [9, p. 60].

However, 'liveness' has a more common use within DMI research, denoting not the method of consuming the performance but rather the causality

[12] http://www.wekinator.org

of the inputs and outputs, or the *perceived agency* of the human performer. Berthaut et al. [23] contribute ideas on this kind of liveness within the NIME context through a study that begins by stating that liveness is 'a problem' within DMI performance, because a DMI offers less visual information than an acoustic instrument, preventing the audience from understanding the music. This study positions this kind of liveness as a kind of perceived *causality*, and the authors suggest that strengthening the causal link between performer action and sound output will increase the perceived sense of 'liveness'.

Auslander engages with this aspect also, and aligns this kind of liveness with a sense of *authenticity* — a performance that involves effort, skill, and is therefore experientially different, and more valuable, than one that is recorded. [9, p. 78]

The DMI literature suggests that it is this kind of causal liveness that is most important to audiences. A study by Marshall et al. [135] measured 'liveness' through emotional response. In this study, students created a DMI and then performed with them, and the audience recorded their emotional response, enjoyment, and their estimation of the performance's liveness. The authors found a positive correlation between enjoyment and liveness, and that 'positive' emotions showed a correlation with liveness.

Like the previous study, Marshall et al. depart from the view — widely held in NIME — that 'hidden' interactions (to borrow a term from Reeves et al. [169]) prevent the audience from establishing a connection between player action and the output, and as such the notion of 'liveness' is positioned as a quality of performer causality as perceivable by the audience.

However, neither study offers a rigorous investigation of the idea of liveness, and rather view it as a type of causality (separate from Auslander's primary identification of liveness as temporal co-presence). This separation is significant, as 'liveness' is used interchangeably to denote these two very different qualities with little examination of the difference between them. Even in laptop performances where there is virtually no perceivable causal link between performer actions and sound outcomes, the audience and performer are still physically co-located in real time; according to Auslander's most basic definition, the performance is 'live' in the sense that the performer is there in person, but not 'authentic', to borrow a word from Auslander, and it is this causal authenticity that is considered to be more impactful.

A notable study by Bown et al. [32] studies precisely this liveness in terms of performer agency. The authors, using audio recordings of laptop performances, asked participants to listen to and rate the performances according to the degree of the following: Familiarity, improvisation, liveness, mistakes, whether the work was pre-conceived, performer involvement, generative elements used, use of other interfaces, and enjoyment.

The relevant finding of this study was that a lack of 'mistakes' — tight timing, aligned beats — was correlated with non-liveness (i.e., not performed by a human), and that perceived performer activity in the recording was correlated with liveness (i.e., performed by a human). Though this study used recordings that respondents listened to through an online interface, the stripping away of context does place these results firmly in the realm of music perception, and suggests that a human — not a computer process — being demonstrably in control of the sound output is important to audiences. Further, this human element is associated with mistakes; a performance that is too 'clean' is viewed as the having lower human input.

This perceived lack of agency has garnered criticism of a wide range of music, not just laptop music and experimental DMI practice. From controversies over lip syncing pop stars to the dismissal of DJs because 'they might as well be checking their email' [32], there seems to be a general expectation among audiences that a 'live' performance (one where the audience and performer are temporally co-present) will also be 'live' (there will be perceivable input from a human agent that the audience considers authentic).

There are, of course, exceptions to this rule. Organs are typically huge instruments in churches whose human agent, the organist, is usually not visible to listeners, and often large parts of the instrument are also hidden in the building. Similarly, many DJ sets rely on laptops and gestures that do not demonstrate their involvement with the music, but DJs and techno music continues to be immensely popular as a live experience.

What separates these examples is the *cultural expectation of the audience of that given musical genre*, part of what Gurevich and Fyans term 'musical ecologies' [87]. For example, no audience of a concert pianist expects that an algorithm may be helping in the playing, and similarly no spectator of an organ concert or techno DJ expects that the player should be displaying their human agency. However, experimental DMI practice, by virtue of retaining the 'formal constraints' of the *Werktreue* of the Western classical tradition [89], also retains the cultural expectation that the musician will produce music via direct manipulations of the instrument, and not doing so disrupts this tacit knowledge. As Gurevich and Fyans state, 'what we know and assume about spectatorship cannot simply be transplanted from the domain of acoustic performances to that of digitally mediated ones and be expected to hold true.' [87]

Even if this expectation is tempered in an experimental DMI context by spectators knowing that there is a computer involved, there is still a question of how an audience can ascertain the way in which that the human performer is making decisions and taking actions that are resulting in an outcome that is masterful, and to what degree this process is human or computer. With neither musical vernacular to rely on against which to judge the output, nor any frame of reference to measure the input of the human performer, the audience is left with a flat space of human manipulations and musical outputs, and without the clues that would allow them to construct meaning.

This also leaves open the question of **virtuosity** in this context. To be virtuosic, according to Dobrian and Koppelman, is to have 'complete mastery of an instrument, such that s/he can call upon all of the capabilities of that instrument at will with relative ease' [62]. Though they specifically differentiate virtuosity in this context from simple 'extravagant displays of extreme speed or dexterity' and acknowledge that 'a computer is capable of playing at speeds much greater than humans, so playing fast notes on a computer instrument is no longer necessarily a display of virtuosity by the performer', there no specific definition of what might constitute virtuosity in this context of DMI performance.

As a result, the carry-over from the Western classical tradition informs audience expectations and notions of virtuosity, yet clashes with the presence of the computer and its potential for automation. This is not an issue confined to DMI performance: Walton, in the context of aesthetic theory, states that 'virtuosity in music ... may replace inspiration and insight' [203], but the computer's ease of performing repetitive tasks heightens this problem considerably.

There does appear to be a confluence of factors that add up to an event horizon where computer influence appears to audiences to overtake that of the human performer. Musician and artist Robert Henke, in reflecting on this effect, remarks:

> *The Uncanny Valley. I spent five years developing and refining my real time laser drawing software for Lumière. Now the results are as complex as technology allows for. I spent a huge amount of time perfecting the sound engine and invested in real time control of sound and image and spatialisation to make every performance unique, and as perfect as possible in a given space. The result: People complain that it is all pre-produced and that there is no interaction with the audience.* [95]

Ultimately, resolving the precise ways in which the presence of a computer inform audience perception of liveness (in terms of human agency) and thereby communicate virtuosity are beyond the scope of this thesis. But, it is important to recognise the profound effects of this disconnect, its results, and the confounding element it presents for audiences of DMI performance.

2.3.4 New directions through limits

The issues above suggest that transparency is essential for audiences to build expectations, but a lack of musical vernacular, unfamiliar and idiosyncratic instruments, and the interference of the computer in the perception of human agency and the additional complexity of computational processes, transparency is exceedingly difficult, if not impossible, to achieve without imposing a new system of vernacular.

But, recent research suggests that another approach is possible. Gurevich and Treviño, in their discussion of an ecological approach to NIME music, raise the relevant issue of 'style' [89]. They define style according to Verplank, calling it 'the ability to perform a prescribed act (e.g. play a melody) in a unique and personal way.'

They argue that the locus of style is in the myriad of personal and artistic choices in real time during a performance, but this definition does suggest the existence of external vernacular from which to deviate in a meaningful way, or at the very least audience expectations that suggest limits to push. But, this is complicated by the nature of this specific musical genre: The limits that are usually imposed by vernacular do not exist because there is no vernacular, and to impose a vernacular would be to nullify the radical experimentation that typifies DMI performance practice.

Gurevich, Stapleton and Bennett later expand this concept, and propose designing instruments with inherent *constraint* as a way of encouraging style. This strategy, instead of addressing this need for limits by proposing the community adopt musical vernacular, instead imposes limits through the design of the interface, thereby preserving the freedom and experimentation inherent in this kind of DMI tradition.

What the addition of constraint adds is boundaries outside which to stray that are apart from the musical or performance context. In a subsequent study, Marquez-Borbon, Gurevich, Fyans and Stapleton [134] found that constrained interfaces (in this case, a box with one button on it) allowed performers to leverage their own musical skill and play in their personal style. Further, they found that 'it takes very little complexity to confuse spectators' and that just one gesture producing two results led to confusion of the gesture-sound relationship; in other words, negatively impacting the instrument's transparency.

A further study by Zappi and McPherson [214] tested this in light of Dix's ideas of appropriation [61], or the idea that users will overcome limits by appropriating other behaviours. This study confirmed Marquez-Borbon et al.'s suggestions that constraint gives rise to style, and added another dimension: After considerable investment of time and experience with performance, musicians who used instruments with one degree of freedom felt there were **more** features left to explore than performers who played instruments with two degrees of freedom. In other words, 'the emergence of diverse and unusual playing styles is a general feature of highly constrained instruments rather than a reaction to one specific instrument design' [214].

Though neither of these publications detail rigorous audience data and are focused primarily on the experience of the performer, these findings are still significant as they demonstrate usefulness of *perceivable limits* in the absence of musical vernacular, and that straying beyond these limits has artistically successful results. This aligns with past assertions within DMI research that the limitations of an instrument are more interesting than its freedoms [47, 175, 131].

Given the suggestion that style exists in the way that performers stray beyond limits, this suggests that understanding audience perception of error may be a rich vein of insight into the kind of limits that are artistically useful in a DMI performance context. I survey the notion of error and its relevance to DMI performance in Section 2.5.

2.4 The influence of HCI on DMI thinking

In the previous section, I summarised the confounding factors inherent in DMIs, concluding that audience-focused inquiry is useful. However, the audience remains under-studied in the DMI domain. This section examines how DMIs are studied, and casts a critical eye over the suitability of current approaches.

2.4.1 NIME: A subset of CHI

Thinking and research in the NIME community is heavily influenced by HCI, for two reasons.

First, DMIs were early examples of computers that are also creative instruments, an intriguing area of inquiry for HCI research. In fact, the research NIME community, where the majority of DMI research is located and where many DMI performers are also located, itself began as a workshop at the ACM CHI conference [166]. This workshop was focused on the question of 'how to better play musical computers' [108], and the approach to 'how' departed from querying the nature of the human/computer relationship.

Second, computer research is largely funded and facilitated by university and government bodies that can afford emerging technology [97, p. 198], particularly that without any immediate commercial application. As a result, this kind of research was typically carried out by engineers and HCI experts, de-emphasising the artistic goals of DMI music performance. (Of course, conferences and communities for computer music such as the International Computer Music Conference (ICMC)[13] existed long before NIME and largely outside academia and HCI research. However, but the influence of HCI paradigms, simply by the virtue of DMIs being computers and where the research has historically taken place, cannot be underestimated.)

2.4.2 Implications of HCI influence

One influence of HCI is that much DMI research is concerned with the performer/instrument relationship. As a result, the audience remains vastly under-studied [11]. From a music perspective this seems counter-intuitive, as

[13]http://icmc.org

this approach privileges one aspect of the performance while leaving behind a plethora of other aspects that are equally influential on audience experience.

But, in the context of HCI, this approach makes sense: HCI research studies the human interacting with a computer, and therefore places the emphasis on the performer/instrument relationship, the locus of the human-computer interaction. (There has been some movement towards considering the audience in HCI study, which I summarise in Section 2.2, but the study of kind of second-degree HCI — watching a human interacting with a computer — is by no means as common or widespread).

Additionally, HCI's history, which spans less than 100 years, has until relatively recently been primarily concerned with task-based interactions and designing ways for computers and humans to interact in every conceivable context: In short, thinking about the computer as a tool that performs a well-defined task, and how best to design it so the human operator most easily completes that task.

This is the core friction between HCI and DMI performance practice. Musical instruments are not tools, in a task-based sense; as Tanaka notes, an instrument is not utilitarian, and is not designed for a single well-defined application. [189] Instead, as Jordà points out, it is precisely the diversity of affordance that offers the player a range of creative possibility [109], and it is the way he or she makes decisions about which affordances to use and how that constitute musical style [88] (a finding that has been further demonstrated [134, 214].) As a result, the task-based HCI approach is not entirely suitable for understanding interaction factors related to DMIs, and disregards, or fails to adequately engage with, the subtler and more nebulous factors at play.

HCI's evaluative goals have also been expressed in DMI research. Evaluating DMIs has been an area of intense scrutiny, and a plurality of frameworks and taxonomies to describe DMIs and evaluate their effectiveness have been developed [27, 93, 99, 154, 185, among others]. This kind of evaluation has never before been a widespread aspect of the creation of traditional instruments, which instead developed through centuries of iterative design and craft practice, the proof of an approach's validity seen in the success of performers and the response of audiences. However, DMIs are expected to be 'proven' to be effective via studies and evaluations.

This is not to suggest that one approach is better than another; indeed, huge amounts of insight have been gained through evaluation in the DMI domain in the space of only a couple of decades. However, this distinction does serve to highlight that artistic goals and the goals of HCI are not uniformly compatible.

One effect of this incompatibility is that DMI evaluations overwhelmingly privilege the perspective of the performer, leaving the audience behind [11]. O'Modhrain [154] asserts that it is important to consider the viewpoints of all stakeholders involved in any DMI (of which the audience is one), but it is notable that the first body of work within NIME that engages directly with the spectator experience is that of Fyans et al. [74, 75, 73, 72], and that the focus within the research community continues to be placed on the experience of the performer.

Though its history is rooted in task-based interactions, HCI's focus has begun to shift in recent years from the task to the experience of the user. As computers have diffused into every aspect of life and our social, emotional, leisure and imaginative lives now are increasingly mediated by devices (a phenomenon that has been termed HCI'S 'Third Wave' [30, 92]), there have been relatively recent pushes to understand these kinds of interactions [98, 77]. Despite this shift, however, this kind of analysis is relatively rare and not as well understood as task-based interactions, likely because it cannot rely on effects that are easily measured and quantified (such as audience perception).

The need for a new approach

The above section details why HCI's existing tools and approaches may not be suitable for evaluating and understanding DMIs. This section extends this line of inquiry to engage with whether DMIs should, as tools of artistic creation, be judged at all, or whether this judging is better left to those who experience the music.

The core incongruency between HCI evaluation approaches and musical outputs is that music's value is not determined by factors far larger than how effectively or efficiently a performer uses an instrument. For example, within the Western classical tradition, that might be how effectively a performer adheres to the score and playing traditions in order to deliver the intent of the composer, and in turn the effectiveness of that composer's artistic statement. These judging frameworks — the elements of *Werktreue* that guide style, the audience's expectations, the emotional communication of the composer, the taste and judgment of the listener — are far outside the player/instrument interaction.

Though this huge number of vast contributing factors makes judging music seem in some respects futile and unknowable, we continue to judge music relentlessly. Both amateurs and experts judge and watch others judge, and the overwhelming popularity of shows such as *X Factor*, the worldwide *Idol* franchise, and *The Voice* demonstrate this as a compelling type of popular entertainment. Despite all this judging, the specific musical tool or method of making the music is rarely discussed, and the opinions of the performers themselves is entirely absent.

Despite this plurality of factors that feed into the evaluation of musical output, DMI evaluations are largely based on performer feedback. They assess a DMI not by its output, but rather if it meets the performer's needs, whether the performer likes playing it, how the performer reacts to it, and so on.

The NIME literature contains a huge number of these evaluations, ranging from studies of the ergonomic aspects of a DMI to assessment of qualitative and quantitative performer feedback [185]. In 2015 Barbosa et al. published a paper exploring what 'evaluation' means to the NIME community [11]. In this paper they counted the evaluations that settled on three perspectives: The performer, the designer, and the audience. They found that evaluations in NIME predominantly focused on the performer's perspective (N=52). Evaluations considering the designer's perspective came in at N=28, and those that considered the audience came in last, with N=20. They suggest that 'the NIME community tends to under-consider the audience in the design of DMIs' (though they do note that these numbers are only from papers reporting an evaluation).

It is also worth noting that Barbosa et al. [11] found that qualitative data in the form of questionnaires and interviews appear to be the preferred method of data gathering within the NIME literature, indicating that our understanding of audience perception in this domain is primarily reliant on reflective reporting of experience, and there seems to be little investigation of what audiences think in real time. (I return to this discussion of methods for audience study in Chapter 3.)

Barbosa's finding of the under-studied audience remains a telling indication that the existing HCI perspective is primed for evaluating the effectiveness of an interface for a given user, but not for that user's audience, and therefore its usefulness may be limited when we want to talk about judging DMIs from an artistic standpoint. Indeed, HCI research, as O'Modhrain [154] points out, often culminates in lists of design principles that represent current knowledge about how a given group of users live, act, work or learn. Acknowledging and departing from that HCI standpoint, O'Mohdrain proposes a methodology for evaluation that considers the multiplicity of stakeholders that exist for any given DMI [154]. This methodology suggests that a DMI has a multiplicity of stakeholders — the composer, the performer, the audience, the manufacturer, and so on — and that this group of stakeholders might have entirely divergent opinions on which aspects of a DMI are the most important. In this way O'Mondrain suggests that the evaluation of a DMI must be more far-reaching, a suggestion that moves DMI evaluation closer to the way music is judged in wider culture. However, it still does presume that the DMI is the thing against which to judge musical output.

These examples are not to suggest that the total of HCI research is not useful, or fails to grasp the important aspects of DMIs. There have been intriguing and useful publications considering what it means to watch someone else interact with a computer such as that contributed by Reeves et al. [169], who suggest that the combination of an interaction with a system and the resulting outcome guides how the spectator understands it.

Despite this, gaps in our understanding of how to judge DMIs and understand what they do are obvious, and indicate that this research is in its infancy. The need for DMI-specific ways of understanding and talking about instruments is needed, as HCI's methods are not widely suitable, and do not produce outcomes that are useful to our knowledge of how to use computers as tools of musical creation.

2.5 A survey of error

The idea of *error* is vastly under-scrutinised in contemporary DMI and HCI contexts. In this section I unpack this notion of error, examining definitions across various disciplines related to DMI research.

2.5.1 Error: Differing definitions, differing effects

Error is a deceptively simple term. Its Latin root is the verb *errare*, meaning to stray; to err is to step outside the accepted way of doing things, to go out of bounds.

Those boundaries present new subtleties. How do we know where these boundaries are? Who determines their location? Can they change? What about the person making the error, and what if they do it in public? Does error exist if the person watching doesn't notice? Is the error in the making, or in the perceiving? Every discipline has its own ideas about how not to do things, and there are a number that inform error in electronic musical performance. I will examine each of these in turn.

2.5.2 Science, computing, communication theory

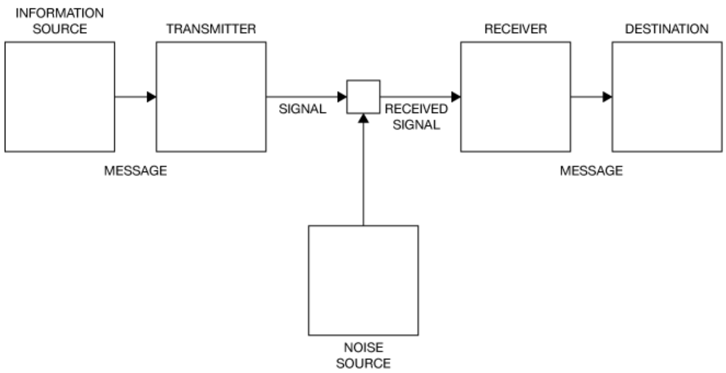
In scientific disciplines there are nuanced types of error (such as discrepancies, judgement errors, and so on), but to be judged an error an action is always compared to an externally-defined and knowable truth. [18] In this way, error is a very straightforward phenomenon. No matter how the error is classified, the effect is always the same: An experiment's results do not accurately represent the truth of the reality of what happened.

This external truth exists in computer science, but the 'truth' against which actions are compared is not the rules of the natural world, but instead the human-made rules of the compiler. In the case of a logical error, an instruction doesn't make sense at all to the compiler and therefore is rejected. In contrast, a run-time error is passed by the compiler but when it is executed it creates unintended results. The error is not in the instruction, but in the *outcomes* of that instruction. This is an important distinction; that an error is not necessarily significant because of what it **is** but rather because of what it **does**.

Communication theory presents us with an intriguing and much more nuanced view of error. The work of Claude Shannon, in particular his seminal paper A Mathematical Theory of Communication [178], is concerned with the 'transparent communication', meaning, in the simplest of interpretations, a signal that is free of noise and arrives at its destination exactly as it was sent from the source.

For Shannon, noise is the error, and enters the message in the channel

Figure 2.17: Shannon's diagram of communication.



phase of communication (Figure 2.17), or the medium through which the message travels. In this sense, error is not a binary state, but rather a cumulative entity: Error, though it impacts with the message, it does not render it useless.

Shannon, too, measures results against an outside 'truth', which is the original message and it came from the transmitter. He represents a transparent communication as source $H(y) = H(x)$, meaning that the output of the communication is exactly equal to the input. Shannon admits, however, that a noiseless communication 'cannot be achieved with any finite encoding process but can be approximated as closely as desired' [179].

To sum up: errors, according to Shannon, are an inextricable part of a transmission, but a transparent communication can be approximated extremely closely even if error cannot be eliminated, and even if errors are noticeable. Even at this noticeable level, error is still not a binary state, and the presence of error does not render a signal useless; a signal that contains noise can still be understood by the receiver until the point at which the signal is entirely eclipsed by noise. What is salient about Shannon's suggestions is that error is not a one-time event, but a cumulative effect, and that the point at which it becomes significant is determined by the receiver of the transmission.

2.5.3 Error in HCI

Research in human-computer interaction (HCI) research is a deep trove of thinking around error. Most of this thinking, however, is concerned with task completion and not of exploration or creative output. Nevertheless, there are several interesting and relevant models to draw upon.

It should be noted that HCI thinking on error has grown out of a tradition of preventing industrial accidents, as the means of industrial production, mass transportation and other large-scale enterprises involved more and more people, and became controlled by humans operating computer interfaces. In HCI research he phrase ‘human error’ is steeped in layers of emotional meaning because, if one of these humans made a mistake, the effects were dramatic and overwhelmingly negative (such as accidents involving aeroplanes and nuclear power plants).

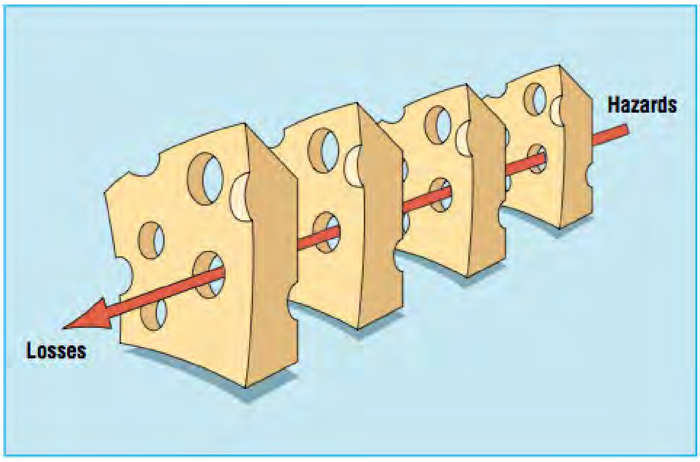
Donald Norman (who, along with being a design expert, specialises in human error in industry) has written extensively on error, and outlines his ideas in his iconic book *The Design of Everyday Things* [152, pp. 162-216]. He states that error is a deviation from ‘appropriate’ behaviour, a term be places in quotes because the ‘appropriate’ behaviour is often unknown at the outset, or only determined after the fact. He also notes that between 75 and 95 percent of industrial accidents are due to human error, but he argues that the human is not the problem; rather, error is the problem of a badly-designed system.

Norman’s intriguing way of framing error places the boundaries of error squarely with the desires of, and processes determined by, the living, human designer(s) of the system. Further, he places the failure to understand human mental processes around interaction with those same living, human designer(s) of the interface (which are sometimes, but not always, the same people). He is careful to continually remind us that this is not a case of hapless human users vs. faceless computer machines, but rather a problem of people not understanding how to design for other people. In this way, he implores us to understand those using systems, and to build systems with better understanding of how humans use things.

Further, Norman presents a model for error which involves two categories of error: *slips* and *mistakes*. Slips are errors committed by the person using a system, due to lapses in memory, physical slips, mix-ups, and so on, such as peeling an orange and placing the fruit, not the peel, in the rubbish.

By contrast, mistakes are more pervasive, more serious, and are the fault of the designer of the system. For example, designing an interface in which all the buttons look the same is a mistake; when a human commits an error by pressing the incorrect one, it is superficial to class this as simple human

Figure 2.18: Reason's 'Swiss cheese' model [168] of how safeguards place a buffer between individual errors, and the irrecoverable state of accident.



error. The error, instead, is not fully understanding human factors at work when pressing buttons, such as fatigue, habit, and other reasons slips occur.

It should also be noted that although Norman's ideas of error tend to be quite binary in that an error is either committed or it isn't, he does admit that errors happen all the time and, like Shannon, hints at the idea of an error event horizon. For Norman, this is the *accident*, the worst possible outcome, the point at which errors are so numerous — or so serious — that whatever process is in motion stops. This is an exceptional circumstance that he explains using Reason's 'Swiss cheese' theory [168], which likens each error to a piece of Swiss cheese, and posits that an accident occurring is like having a hole on all the pieces line up at once, forming a path through all the layers. In this model, safeguards (such as the ability to undo an erroneous gesture) provide further buffering layers between the single error and the accident event horizon, the metaphor being that it is harder to line up a hole through several pieces of cheese as opposed to one or two (Figure 2.2).

Norman's work evolved in parallel to that of Jens Rasmussen [167], who developed the Skills/Rules/Knowledge (SRK) classification of error. Rasmussen engages fully with the fact that human behaviour often evolves and is modified as it heads towards a goal, a phenomenon termed teleology [172]. Instead of seeing all human tasks as discrete actions, Rasmussen instead sees them as a complex sequence of composed events: *'In general, human activities can be considered as a sequence of such skilled acts or activities composed* *for the actual occasion. The flexibility of skilled performance is due to the ability to compose, from a large repertoire of automated subroutines, the sets suited for specific purposes.'* [167, p. 259]

Error and HCI's Third Wave

As computing has found its way into every facet of our lives, the experience of the user has become a central point of HCI study and discussion [211]. As interfaces have become ubiquitous and their applications have expanded to include a broader and more personal range of interactions, evolving far beyond task-based interactions in the workplace [152]. Computers are now the locus of experiences that have no particular task, or a task that may only arise through doing, such as social, emotional, and leisure experiences. These experiences are ambiguous and as such the interfaces built for them benefit from a certain degree of ambiguity. Gaver et al. [77] propose that this ambiguity in interfaces is a 'resource for design' that, instead of leading users through a task, instead provides a space of possibility for interpretation. Further, Sengers and Gaver [177] assert that HCI 'can and should systematically recognize, design for, and evaluate with a more nuanced view of interpretation in which multiple, perhaps competing interpretations can co-exist.'

Among these ambiguous interactions are those relating to ones. A 'third wave' of HCI, which states that thinking is abstract and that we make meaning by, for example, doing things, manipulating interfaces and using expressive gestures [92], moves away from the goal-oriented idea of interaction, and frameworks such as Benford et al's classification of *expected*, *sensed* and *desired* gestures [22] form a way of thinking about interactions that may not have a specific goal.

Of course, moving the discourse from task-based interaction to include more fluid uses of interfaces is a positive trend that expands the discussion around HCI to include a great many more nuanced and emerging ways of using computers, particularly for creative means. However, absent from these analyses and frameworks is full engagement with what it means to do something wrong, and what role that plays in the creative output. If there is no 'appropriate' behaviour, where does that leave error events? Even more interestingly, where does that leave successful ones?

Benford [21] analyses musical performance directly using this expected

/ sensed / desired model. Benford posits that part of the challenge of interacting with these systems is that their processes are often invisible; the computer is doing the 'work', as it were, and it seems to the performer, and to the audience, that the gesture has had no effect. Expanding further on this idea, Reeves et al. presents a framework for classifying performance gestures — secretive, magical, expressive or suspenseful — as a means of understanding the *effects* of certain performer gestures. (I return to the discussion of this framework in Chapter 3.)

These frameworks are extremely useful, particularly the latter taxonomy, which aims to classify gestures into the four categories and provide some insight on how to design interfaces for the experience of the audience watching their use. This article is also some of the only writing that hints at error being a creative opportunity; Benford does suggest that performers wishing to be more exploratory may 'experiment with novel or extreme ways of playing the instrument in which they must push themselves into unusual positions and actions to create particular sounds, which could lend an interesting dynamic to their performance' [21, p. 52]. Exploratory interactions are risky gestures, because they create the opportunity for error; Benford identifies this risk as possibly being a dynamic performance element.

2.5.4 Error in music

In scholarly literature on music, writing that engages directly with the notion of error is conspicuous by its absence. It seems to be taken for granted that error is completely unwanted and should be avoided.

In terms of the error boundaries, the Western classical music tradition has boundaries that are rigid and quantifiable, such as score, tempo, pitch (see Section 2.2 for a discussion of *Werktreue*). This provides a convenient 'truth', or true version of a given work, against which to measure output. Some cognitive scientists make good use of this in order to measure musical behavior [162], and to produce mathematical models of performance [58, p. 352].

Though *Werktreue* is a pervasive idea and it is tempting to classify music based on what follows and what resists it, there are plenty of exceptions and examples of friction. Andy Hamilton wrote of the 'aesthetics of imperfection', or musical practices that fall outside, around, or directly opposed to this aesthetic hegemony, particularly the practice of musical improvisation. He differentiates the two modes of practice thusly: 'The aesthetics of imperfection thus focuses on the moment or event of performance, while its rival emphasizes the timelessness of the work.' [90, p. 170]

There is also plenty of discourse around the conforming with, and resistance to, the concept of *Werktreue*, most of which is outside the scope of this paper. However, it is worth mentioning that though error is seen as singularly undesirable in musical discourse, some errors serve to add to music by supplying musical style. Barolsky, when examining Moiseiwitsch's (recorded) performance of Chopin's E Minor Prelude [13], calls his deviations from the score 'idiosyncratic' and argues that these deviations serve to illuminate the Prelude's formal use of texture. It should be noted that the 'errors' to which Barolsky refers are intentional on the part of Moiseiwitsch, but are nevertheless deviations from the score, and instead of lowering the value of the music they serve to show us more. After presenting a case for error's usefulness, however, Barolsky, curiously provides a caveat: 'This approach serves not to defend idiosyncratic and controversial interpretations. Instead I argue that as performers, analysts, and critics, we can learn much from an unexpected turn of phrase and even from the occasional error' and states that mistakes 'may serve to distract us from the musical whole', so in this analysis musical vernacular still survives intact.

Polished to death: Auto-Tune

In this section I examine the effects of Auto-Tune, the plugin for Digital Audio Workstations (DAWs) that corrects pitch errors in vocal tracks.

However, the location and flexibility of these boundaries are what determine if an action has crossed into *errare*, and sometimes actions or features get mis-categorised. An example of mis-identifying error because of an incorrect or incomplete idea of what error is the phenomenon of Auto-Tune. Auto-Tune was developed in the 1990s by Dr Andy Hildebrand, who applied his knowledge of auto-correlation (which he developed working in the oil and gas industry) to recorded music. Given an input (in this case, a vocal track), Auto-Tune adjusted the input to the nearest acceptable pitch (determined by user settings). It was released in 1997 as both a software plug-in and as a piece of rack-mounted hardware. [59]

In 1998, Auto-Tune was used in the wildly popular Cher song, 'Believe', and the effect on Cher's voice (colloquially known as sounding like it was being 'sung through a fan') caused a sensation. In the nearly two decades since, Auto-Tune has been applied to a huge number of pop songs, and has contributed significantly to the aesthetic of 'perfection' in contemporary pop music. Author Jennifer Blackwell likens this perfection aesthetic to the unrealistic and unattainable standards of female beauty set by the fashion industry [29]. This standard certainly is unattainable, and in some ways has devalued actual human performances because they're full of stylistic diversions. For example, Aretha Franklin's performance at the Obama inauguration was criticised for being off-key, and that this criticism stems directly from the fact that it was actual singing and not processed, Auto-Tune-perfected vocals [197]).

Despite its widespread and liberal use, the practice of using Auto-Tune has come under considerable criticism. This criticism generally centres around the fact that Auto-Tune processed vocals are devoid of stylistic features and personality. David Byrne writes of 'the uncanny perfection' of Auto-tune being facile, obvious, and 'ultimately boring' [38, p. 132] As NPR music critic Tom Moon remarked, '[T]here's a certain quality that you get from him that no one else in the world has, and the minute that you put him on the grid and align him, as happens with auto-tune, you're in a different business. [S]uddenly, something that's essential about Neil Young, something about his 'Neil Young-ness' is taken away.' [59, p. 5]

This criticism suggests that perhaps within these 'errors' are the marks of human performers, unique fingerprints of musical style; performers from Frank Sinatra to Bob Dylan to Aphex Twin are known for their singular ways of departing from the norm. Therefore, to imbed something such as Auto-Tune deep into popular music is to impose a newer and ever more rigid *Werktreue.*

The rightness or wrongness of correcting vocal performances beyond recognition is an aesthetic argument that is beyond the scope of this research. However, we can deduce this: the literature around Auto-Tune suggests that identifying these 'errors' and correcting them has not had the intended effect of somehow improving music; on the contrary, this 'correcting' process has brought with it subtler problems, and what value has been added is wildly debatable. Considering the effects and critiques of Auto-Tune, it certainly seems possible that the working idea of error — *errare* in its truest sense, straying from the correct pitch frequency — is reductive and simplistic, and that this attempt to fix the 'problem' has deeply misidentified its cause.

In short: Error is not so straightforward, so quantifiable, so rigidly consistent across performers and genres. Auto-Tune has been applied liberally and widely as some sort of sonic panacea to perfect and correct music, but it seems that what we have learned is that error isn't what we think it is, and that a more useful definition is needed.

2.5.5 Philosophical notions of error

> When you invent the ship, you also invent the shipwreck; when you invent the plane you also invent the plane crash; and when you invent electricity, you invent electrocution... Every technology carries its own negativity, which is invented at the same time as technical progress. [201, p. 89]

Along with the definitions of error from various research contexts, there are also philosophical interpretations of this term that are useful.

Cultural theorist Paul Virilio is the most prominent thinker in this space. Where HCI error theory tends to anticipate disaster on the level of a plane crash, Virilio anticipates utter catastrophe at a global scale. The root of our careening towards disaster, according to Virilio, is *speed* — our existence is defined by the 'twin phenomena of immediacy and of instantaneity' [200] which, he asserts, have created an acceleration towards disaster that we have no hope of avoiding.

This is a definition of error that is out of proportion with errors in DMI performance, but Virilio's identification of the problem with computerised acceleration is relevant. The introduction of the computer into the musical instrument paradigm has brought about instantaneity, and is the root of the confounding factors we find ourselves trying to address. The computer has brought us control dislocation, by making it possible to a huge range of sounds with the flick of a finger, and breaking down the gesture-to-sound relationship that has existed for millennia. The computer has resulted in the problem of transparency, by allowing the creation of extremely complex input-to-output mappings. Electricity may have been labour saving in the past, but computation is labour-saving accelerated — never before has so much been possible by doing so little.

In this way, the acceleration brought about by technology is at the root of our most confounding concerns. But, despite his warnings of doom, Virilio is not anti-technology; instead, he 'resists the mythologizing of technology, and he does so because it is precisely by idealizing technology that we come to unthinkingly embrace it' [80, p. 102] — in other words, Virilio cautions against slavish worship of technology as a panacea. Virilio reminds us that all technology contains the potential for error, on any scale, but he does not maintain that technological accidents are only negative; in fact, the technological accident 'can reveal something absolutely necessary to knowledge' [130, p. 63]. The accident — or in the context of this thesis, the error — serves the function of revealing a system's entropy, and it is through recognising this entropy that we come to understand its function.

In the DMI research community, Kim Cascone offers a perspective on the 'post-digital' in computer music [45]. He uses the term 'post-digital' because 'the revolutionary period of the digital information age has surely passed', and what we are left with is the ability to zero in on and capitalise upon errors in digital systems. Quoting David Zicarelli, who said in 1999 that 'failure tends to be far more interesting to the audience than success', Cascone not only establishes a way to use our technological error for creative ends, but also proposes that there is plenty to learn through interesting error.

In the two decades since Cascone's work, glitch has gained significant momentum as a genre in art. Glitch, which embraces technological errors as an artistic opportunity, has flourished. While many artistic disciplines seek to avoid error, glitch adopts it.

Glitch artist and theorist Rosa Menkman, in 2011, developed a Glitch Studies Manifesto. The first tenet is a challenge to Claude Shannon:

> **1. The dominant, continuing search for a noiseless channel has been — and will always be — no more than a regrettable, ill-fated dogma.**
>
> Acknowledge that although the constant search for complete transparency brings newer, 'better' media, every one of these improved techniques will always possess their own inherent fingerprints of imperfection. [139, p. 11]

Menkman urges us to accept that everything is imperfect; we may make improvements, but errors will always be present. Instead of worrying about perfection, glitch art moves swiftly on, and 'deals with the digital dimension of error, accident and disaster' [139, p. 32].

2.6 Conclusions and research questions

In this chapter I map the theoretical context for this thesis. In Section 2.1 I explore the history of DMI performance, and trace its roots from the 20th century avant-garde to the present day. In this section I also provide an overview of the influences on experimental instruments and musical practices from modernist practices, and describe the influence of electricity as a labour-saving technology on the development of DMIs — the origins of one of the DMI's most confounding characteristics. Further, I provide an overview of contemporary DMI performance, demonstrating the width and breadth of this performance practice, and define the scope of DMI within this work.

Next, in 2.2 I describe Lydia Goehr's theory of *Werktreue* and how it applies to DMI performance. I make an important distinction between *Werktreue* and musical vernacular, as it is the latter that DMI performance rejects while taking advantage of many aspects of the former. Additionally, I detail relevant theories of audience perception of music and mental models, concluding that these tend to rely on musical vernacular that is absent DMI performance, and propose that Dervin's sense making methodology may have applications here.

Turning specifically to DMIs, I then describe the confounding factors for audiences in Section 2.3, focusing on the three notions of transparency, familiarity, and agency, detailing the contemporaneous knowledge within the NIME and DMI research literature and identifying areas where insight is needed. Given the demonstrated usefulness of constraint in the design of DMIs, I hypothesise that error may be a locus of insight into the audience experience.

In Section 2.4 I describe the evolution of DMI development and research that has taken place in the HCI research context, and how this has contributed to the popularisation of certain ideas and approaches that do not consider the audience. In this section I survey current ideas related to DMI research and critically evaluate how these can be applied to the understanding of music as an artistic practice, and where these approaches may cease to be useful. I suggest that a new methodology, one that can more fully considers the audience experience and does not assume that the performer/instrument relationship is of utmost importance, is needed.

Finally, through Section 2.5 I examine the notion of error. I consider ideas from a range of disciplines related to DMI research, thereby demonstrating that though it is often viewed as a binary term it contains a high degree of nuance. I examine error within the context of HCI (both past and present paradigms) as well as within the context of music, both of which consider error to be undesirable. Through the example of Auto-Tune I demonstrate that simply removing 'errors' does not necessarily improve music, and that not only might its negative reputation be unwarranted, but reinforcing that error may in fact be the locus of style in a context that lacks vernacular, and therefore is a potential source of insight.

2.6.1 Research questions

This survey of existing knowledge in the DMI research and related domains suggests a trajectory through these interconnected issues.

This trajectory lies across two veins of inquiry. Within this first trajectory, there are three questions that separate issues so they can be examined through the three studies presented in this thesis:

- **Question 1:**
  With no external stylistic frame of reference, how does the audience perceive error in DMI performance, and how does this affect enjoyment?

  As demonstrated in Section 2.1 and 2.2, DMI music resists many of the rules of musical style, and that the origins of this rule-breaking has its roots in 20th century modernism.

  In Section 2.5 I demonstrate that error may be a locus of insight into audience perception. By viewing audience enjoyment in the context of error (and vice-versa), this has the potential to lend dimension to our understanding of audience perception of DMI performance.

- **Question 2:**
  What is the impact of visible risk on audience perception of DMI performance?

Related to RQ1 above, this question serves to determine whether, if we can understand something about the audience perception of error, whether error is a necessary condition with interesting outcomes.

- **Question 3:**
  Can the physical design of a DMI affect the performative outcomes?

  If indeed error is an important factor in audience perception, then this has implications for DMI design and the role of the physical interface, and there is potential to explore what aspects of DMI design may affect these.

The second vein of inquiry confronts the need, demonstrated in Section 2.2, 2.3 and 2.4, for a new way of studying audiences in this realm:

- **Question 4:**
  How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

Post-hoc data is liberally employed in DMI research, but lacks moment-to-moment granularity; real-time data is employed in HCI research but lacks the context required to infer conclusions about audience response. In Chapter 3 I detail the way in which these questions are explored over the three studies in this thesis. In this next chapter I also describe how this research has been designed in order to fully consider the audience, as well as the factors affecting these design decisions and how they were navigated.

Chapter 3

Methodology

This chapter describes the methodological approach I took to studying live audiences. In it I describe my methodological choices, and the theoretical underpinnings of these.

I chose to carry out this work in a live context, despite the logistical challenge of live audience studies. Section 3.1 describes why this approach was necessary, by defining 'live' experience as distinct from recorded performance footage.

The experimental setup was consistent across all three studies. Section 3.2 explores the rationale for and influences on the decisions made in its design. Here I also describe how these decisions contributed to negotiating the aspects of live experience identified in the previous section.

Choosing to carry this research out in a live context brings with it a great deal of complexity. Section 3.3 identifies these complexities that are inherent in the live music experience, and explains how I navigated these. I discuss where I preserved complexity in order to privilege the important and useful aspects of live performance, and where complexity was controlled in order to produce meaningful results.

I consciously chose to explore error and its relation to the positive notion of 'enjoyment' in order to query the nature of error as it is experienced by audiences. In Section 3.4] I discuss what these terms mean severally, and how they relate to and inform one another.

A central contribution of this thesis is the combined approach to audience data. In Section 3.5 I discuss existing methods of gathering real-time and post-hoc data. I detail the advantages and disadvantages of each and how these data types differ and inform one another, thereby demonstrating the value of a combined approach. I then explain how I performed data gathering of both the post-hoc and real-time feedback.

Following on from this description of how the data was gathered, I then describe my approach to how the data was treated and analysed in Section 3.6. I also provide details of my quantitative, qualitative, and real-time analysis. Additionally, this section details the post-structuralist approach I took to this work, and how this informed my research goals.

Finally, in Section 3.7 I describe how each study contributes to the central research questions described in Chapter 2. Each study approaches these questions by unpacking some underlying confounding factors of DMI performance, and this section details the goals of each study.

3.1 Considerations in the study of live audiences

> *The 'audience' is equally and simultaneously identifiable and elusive, imaginable and unpredictable, and enduringly fascinating for all those reasons.*[36]

This section details factors that are important for audiences of musical performance, namely the qualities of liveness and multimodality that are intrinsic to music performance. Further, this section examines the challenges presented by doing this research in a live setting, and why negotiating these sources of complexity, as opposed to simply eliminating them, is necessary for this work.

3.1.1 The importance of liveness

In Section 2.3.3 I described a definition of liveness that is specific to DMI research, which pertains to perceived human agency, or the audience's ability to see and understand the human (as opposed to computer) involvement in the music-making process. This differs from the use of the word 'liveness' in the context of the live audience. For the latter notion of liveness, I looked to Philip Auslander's definition of 'live performance' as 'the kind of performance in which the performers and the audience are both physically and temporally co-present to one another.' [9, p. 60]. A full investigation of how Auslander describes what it means to be 'live' is beyond the scope of this work, but it is important to note that he draws a clear distinction between 'live' (in person) and 'mediatized' (recorded or broadcast) performance.

Philosopher Peggy Phelan agrees that recording a performance changes it into something else. She states that this is because there is meaning encoded within physical and temporal co-presence:

> *Only life is in the present. Performance cannot be saved, recorded, documented, or otherwise participate in the circulation of representations of representations: once it does so, it becomes something other than performance. To the degree that performance attempts to enter the economy of reproduction, it betrays and lessens the promise of its own ontology ... Performances independence from mass reproduction, technologically, economically, and linguistically, its greatest strength* [163, p. 146]

According to Auslander and Phelan, the live experience differs significantly from recorded (or, to use Auslander's terminology, 'mediatized') experience. As this thesis aims to understand audience perception of live performance, it was essential to do this research within a live context where the audience and performer are co-located, in order to include the aspects of liveness that impact the audience. As such, this investigation recognises live performance as a distinct and separate entity from recorded footage.

This commitment to the live context does introduce several challenges and risks to this research, namely:

- Live events are logistically complex, and difficult to plan and execute
- The audience size is limited to those who are there in person for the experience
- It requires the participation of musicians who have to be recruited
- Because of these factors, there is the inherent risk that any experiment conducted in these conditions may not produce useful data or observable results.

Using recorded performances would be an effective way to mitigate these. For instance, conducting this research online using video footage would deliver a consistent performance to each audience member, and there would be no upper limit on the audience numbers. However, doing this research online would disregard the live, in-person context that supports an essential aspect of music performance that is profoundly influential on the perception of audiences: Multimodality.

3.1.2 The impact of multimodality

> *A person does not hear sound only through the ears; he hears sound through every pore of his body.* [117]

The term **multimodality** has its roots in linguistic and semiotic literature, and 'is used to highlight that people use multiple means of making meaning' [106, p. 2].

As described in Chapter 2 Section 2.3, the importance of the visual components of a musical performance for establishing performer agency (which may include, for example, body movement [202]) are well established. Though some visual aspects may be preserved through video, this does not include the sensation of 'being there' — the quality of light, the experience of sitting as an audience, the atmosphere in the room, the social dynamics, and so on are not reliably transmittable through simple video recording. As Paine states, musical performance 'operates from within complex traditions of culture, musical design, and performance technique' [158], and this complex confluence of factors contributes to the live, multimodal experience.

It is important to highlight why I choose the term 'multimodality' to describe the musical context, and not 'immersion'. There is crossover: In *The Art of Seeing* [49], Csikszentmihalyi and Robinson examine the aestehtic experience of art as a specific kind of 'flow' experience that is characterised by total focus, concentration, and attention on what is happening before them. They describe the aesthetic experience as the moment 'when information coming from the artwork ... fuses with information in the viewer's memory' [49, p. 18]. In this way, the live music experience has features of immersion, but this term implies passive consumption, and does not highlight the meaning-making process present in multimodality. In the case of music, a considerable portion of the 'information' from this immersive experience is processed multimodally, through visual, aural, and sensual modes of perception, and include environmental, social, temporal, and spatial factors. Without live presence, these channels of communication cannot be perceived in the same multimodal way.

It is important to note that although music is known to be a multimodal experience, 'psychological research has ignored non-acoustic aspects of performance, considering them as extraneous and not essential to the music' [193]. In live audience research within the NIME community, this seems to be the case. For example, the live audience studies by Fyans et al. [72, 74, 75] used video footage of performances to draw conclusions about the audience experience. This is not to cast doubt on their findings or methods; their research questions were based around the ways in which audiences understood gestures and visual aspects of performance, and for those purposes of that particular study video footage may be suitable. However, it is important not to conflate, or equate, the 'live' experience with that of a recording.

A NIME study by Lai and Bovermann [124] supports the notion that the performance space is important to audiences. They describe a study using qualitative audience data that found the performance space to be an 'important element' in the experience of spectators because of its sonic qualities, so much so that they cite the performance space as a prominent consideratio n in the production of live performance. Since the audience does recognise the performance environment as an important factor, this suggests that results from studies such as those from Fyans et al. are not directly applicable to questions of live audience experience, because they do not retain the essential quality of multimodality within their inquiry.

3.2 The experimental setup

Each of the 3 studies described in this thesis took place in the context of an evening concert. In this section I discuss the components of my experimental setup, and the reasoning behind these. Further, I discuss how these decisions supported my emphasis on the live experience, and preserved the live and multimodal aspects of music performance.

3.2.1 The performance venue and audience

Each of the 3 studies took place in the Film and Drama Studio in Arts Two, Queen Mary University of London (see Figure 3.1).

Figure 3.1: Photos of the Film and Drama Studio in Arts Two, Queen Mary University of London

The choice of this space was deliberate, as the space is a purpose-built concert venue. It has 65 seats facing a floor-level stage, and has sound and lighting available that clearly delineate the performer from the audience. In this way, the venue retained the 'formal constraints' as identified by Goehr [83, p. 24] that are common of this kind of DMI performance, thus reinforcing the performance frame.

The **performance frame** of these concerts was a primary consideration. First coined by Bateson [14], a frame is a cognitive context with rules and behaviour, and these rules directly influence the interpretations of communication and events happening within that frame. In Western classical music tradition, for instance, this rules of the performance frame are well established: The audience sits in seats that are in rows, is quiet during the performance, does not intervene, and applauds at the end; booing expresses ultimate displeasure; one does not leave until it's over.

As explored in Chapter 2, the work of John Cage experimented liberally with musical vernacular but retained what Goehr calls 'formal constraints' of Western classical music, and the conventions of the performance frame in this context are part of these constraints. Though plenty of practitioners are exploring and experimenting with the way that DMIs, and particularly ubiquitous mobile technology, can interfere with, play against, and disrupt the performance frame [189, 78, 94, 33], these efforts tend to use the established frame and its rules as a point of departure and a constraint against which to push, suggesting that the performance frame of the Western classical tradition is deeply established and influential.

Sheridan extended this notion of the performance frame specifically to audiences of digital live art through the concept of **wittingness** [182, 181, 180], which she defines as 'an individual's or group of people's knowledge or awareness of the performance frame'. Further, Sheridan identifies the audience within this performance frame, asserting that 'the individual (or group) has accepted by choice or without reluctance to interact (or to not interact), and therefore they have an established knowledge about the performance frame' [181].

This is an important distinction in this context. DMI performance is by definition rule-breaking and experimental, so it cannot be taken for granted that any live audience understands the rules and communication of the context they are in. In this way, the pervasive nature of the performance frame of Western classical music that DMI performance adopts is an advantage, precisely because its rules are well known and well established. The performance venue supports this: It has rows of seats, a stage area, lighting that emphasises the performer and de-emphasises the audience, and so on. In this way, a strong performance frame is established. Additionally, the audience was composed of individuals who chose to attend these events based on the event being promoted as 'an evening concert', so it can be reasonably assumed that the audience was, by Sheridan's definition, witting, and understood and was aware of the performance frame and its conventions.

Using the same concert venue also created a consistent baseline of multimodality over all three studies. Though the precise actions and content of the performances were not the same, the surrounding aspects — such as the stage lighting, the acoustic response of the room, the cultural context — were kept consistent.

It is important to note that these same multimodal factors (lighting, acoustic response, and so on) could have been kept just as consistent if the concert was held in a lab setting, and arguably could have been controlled to an even greater degree. However, situating these in a lab setting changes the performance frame, because it changes the cultural and social dynamics: If the venue does not align with audience expectations of a venue for musical performance, it will potentially introduce an amount of disruption to the audience's perception and expectation of these social and cultural dynamics. By placing these concerts with a space designed for performance and consistent with audience expectations of performance, the performance frame is preserved as much as possible.

3.2.2 The event

Each of the studies was advertised as an evening concert, and was promoted through email lists and social media channels.

It is important to note that in all promotions the existence of a study was explicitly stated, and it was also stated that attendees were free to simply enjoy the music if they didn't want to participate. This indication of a study could potentially disrupt the performance frame in some respects, as scientific data gathering is not a usual event in this cognitive context. But, it can be argued that this transparency established an extension of this known performance frame in advance, thereby managing audience expectations of the context into which they were entering.

Each study was composed of three or more 'acts': an opening act, the act containing the study, and a closing act. This served to reinforce the performance frame, and ensure the audience was settled within it before the study started. Each concert followed this structure:

1. The audience was seated and given a program detailing the performances they were about to see.

2. Before the first act a member of the lab gave a short introduction, welcoming the attendees.

3. The first act(s) played.

4. Before the study performance, I introduced the audience to the study, and explaining the purpose and describing how it worked. Audiences were then given survey books and pens, and briefed on Metrix, the real-time data collection system (fully described in Chapter 4) through a 2-minute video. I then answered any questions.

5. The study performance took place. Though this took different forms over the 3 experiments, audiences gave real-time feedback during the

event, and filled out a short survey straight after. They also filled out a longer survey at the end.

6. The survey booklets were collected, and the final act(s) played.

The structure of these events meant that the frame of the evening concert remained as intact as possible, while still acknowledging the presence of a study.

A note on observation

Ethnography has long struggled with the notion that any phenomenon that is studied changes by virtue of being observed. It is not my claim that this methodology renders these concerns nullified, or that all of these complications have been addressed. I also accept that any concert held for the purposes of scientific inquiry, as opposed to cultural reasons, can be considered, in Baudrillardian terms, a simulacra [15, p. 12][1].

A full discussion of the construction of 'authentic' experience, cultural simulation, and how it impacts audiences is beyond the scope of this paper. Still, it is important to note that the purpose of inquiry does not in itself render the simulacra nothing more than an ersatz copy of a ghost of a cultural event; the most important factor is the audience's understanding of the performance frame and its inherent communication and rules, not whether or not the concert could be objectively considered to be an 'authentic' cultural happening. Peterson [161] argues that an event's perceived authenticity is not a function of the 'realness' of its component parts, but rather a socially constructed value; by this token, the experimental context fulfils its role if the audience considers it acceptable.

Still, it is vital to highlight that the purpose of these studies was not to create a seamless representation of a concert and assess that effective-

[1]The term 'simulacra' refers to a copy for which there is no original, which is certainly true of musical events — there are cultural signifiers that might communicate 'concert', but there is no original event to which the term refers. The reference cited is Baudrillard's discussion of simulacra in *Simulacra and Simulation* where he uses Disneyworld as an example of a simulacra. Though this might seem a cynical and overwrought comparison (especially in the terms in which Baudrillard describes it), it is applicable because music is laid heavy with cultural influence: We all experience music, all the time and in all aspects of culture, and therefore everyone carries with them expectations of this experience, further weighed down with cultural meaning. For this reason, a concert for scientific purposes is perhaps not a simulacra, but a simulacra of a simulacra — a representation of a 'real' concert, which is itself a simulacra.

ness, but rather to conduct this research while mindful of the performance frame and the social content within it, and taking steps to ensure that the legitimacy of this frame isn't disrupted simply by being a study.

3.2.3 The music

The following section contains a discussion of the implications of including musicians and instruments in these studies. In this section, I describe the reasons for including other musicians in these events.

Each study event was organised on a theme. These were:

1. Innovative Interfaces
2. Beep/Bang: An evening of experimental percussion
3. Music from the Augmented Instruments Lab

Organising around an event not only kept cohesion within the concert, thereby reinforcing the performance frame, but also served to guide audience expectations of what the performance would entail.

The study act for all three experiments involved professional musicians playing music they had composed on DMIs (for Study 1 these DMIs were of their own creation, and I designed and produced the instruments used in Study 2 and 3). I discuss the rationale for using professional musicians and music they had composed in the following section.

3.3 Navigating complexity

> *Music is understood as a dynamical complex of interacting situated embodied behaviours. These behaviours may be physical or virtual, composed or emergent, or of a time scale such that they figure as constraints or constructs. All interact in the same space by a process of mutual modelling, re-description, and emergent restructuring.* [102]

Complexity is an inherent part of the live music experience. All aspects of music performance — the music developing over time, the place in which it happens, the social dynamics, the audience and all the individuals within it — are intricate nests of complex factors. This section explores the complexities of a live experience, and where these complexities could be useful, and when it was appropriate to control them.

3.3.1 The importance of complexity

The goal of this work is to understand how audiences perceive DMI performance in order to gain insight about how to build instruments with which performers can deliver enjoyable experience for audiences, while also accounting for the experimental nature of DMI performance. This goal is complicated by the fact that this research is as much about how humans can interact with a computer as it is about how humans understand music, and illustrates the friction at the intersection of the scientific and the artistic domains.

In the scientific domain the goal is clear results. In the design of any scientific experiment it is necessary to control independent variables in order to measure their effects. Controlling independent variables is necessary for results to be reproducible, and for conclusions to be credible. Video footage of performance, for example, can control for variations in the way audiences see a performer, as each viewer gets the same viewpoint and lighting, and video reproduction means there can be no variation.

However, this is counter to music's artistic goals: In the making of art it is precisely the presence of the complexity brought by this kind of independent variable that produces the intangible qualities of liveness and multimodality and make up the landscape of musical possibility in which the performer operates in a live capacity. There exists considerable literature on the existence of complexity in music and its interrelated and shifting components, known as the *musical ecology* [102, 206, 53].

It is important to note that currently no applicable methodology exists for the study of live audiences within the HCI and DMI literature. Simply put, this friction between scientific enquiry (which relies on clarity) and musical ecology (which relies on this complexity) has not yet been resolved. In this course of this work it was important not to privilege one domain over the other, and effectively managing this complexity was key to both creating a performance environment that fully engaged with the important aspects of the musical ecology, while also controlling complexity in order to produce useful results.

This section serves to detail how I navigated this complexity, and how complexity was encouraged in some aspects and controlled in others.

3.3.2 Encouraging and controlling complexity

Each of the 3 studies described in this thesis involved DMIs, played by professional musicians. For Study 1 these were DMIs created by the musicians, but for Studies 2 and 3 I designed and produced the instruments for specific investigative purposes.

It would in some ways be more straightforward to use simpler interactions to test audience perception, and ensure that all audiences saw exactly the same interaction. For instance, a MIDI drum pad instead of a percussion instrument would be not have the complexities of DMI mapping, and a video recording of the performance would mean that all audiences saw the exact same interaction. Though this approach seems to address experimental problems, simplifying the interaction in this way would not be consistent with what a DMI performance is, and disregards the multimodality inherent in the musical ecology. Because of this, it would be questionable whether any findings could really be applied to live musical performance, and not only to the specific controlled setting in which the study took place. For this reason, these studies incorporated complexity in ways that supported this ecology. At the same time, constraint was applied to aspects that in ways that meant that these complexities would be not be entirely unbounded. These aspects were:

The concert context. All studies took place as part of an evening concert. Although this was risky as it would be very complex to re-run the experiment, and although this limited the audience to those in the room, it did engage with the aspects of liveness and multimodality that are central to the experience of music in a live context.

Despite these complexities, the performance context was consistent across all three studies, as each took place in the same venue and all were presented as an evening concert. This removes some of the complexity of context - although the audience and musicians are different, it is easier to compare across these three studies where the performance frame is kept consistent rather than across, for example, a performance that takes place in a concert hall vs a performance that takes place at an outdoor festival.

The DMIs were designed for performance. In all studies, the musicians played DMIs that were designed for the purposes of musical expression. Though the DMIs produced for Studies 2 and 3 may have been designed with investigative purposes in mind, they were, primarily, musical instruments designed for musical expression, as opposed to tools that were designed to test a specific aspect of audience perception. These DMIs had mapping of input to output characteristic of this kind of instrument, care and attention was paid to the sound design and quality of production.

Certain selective criteria was applied to the type of DMI used, and for Studies 2 and 3 this causal relationship was a primary consideration. For the Study 1, I sought musicians whose novel interfaces could play in both a conventional and experimental style. For Studies 2 and 3, I designed and produced the DMIs, and chose to make percussion instruments. This was not only because of my personal interest in percussion, but also because the cause-and-effect relationship of percussion already exists and is known to audiences, eliminating some of the issue of control dislocation.

Still, I did not want these instruments to be reductive tools, and the interface designs were carefully considered. The design of the interfaces encouraged the application of the musician's personal playing style through the application of constraint [88] (this is explored more deeply in the instrument design discussions in Chapters 7 and 8). Constraint was used to produce interfaces that would present creative affordance, but would also encourage playing gestures that privilege the cause-and-effect relationship.

The basic barrier to understanding an instrument is this notion of 'control dislocation' [141], which was the tradeoff for a new world of musical possibility. Paine identifies this, saying that DMIs 'heralded the dislocation of the excitation, sonification mechanism, dissolving the embodied relationship musicians previously enjoyed with their instruments while simultaneously introducing a range of possibilities that defy the limits of the human body, raising questions about the role of gesture in musical performance and the value of haptics in successful musical instruments.' [158] Though the question of transparency (in the sense of what it is, how it's achieved, and how much it matters to audiences) is quite nuanced, addressing this basic issue of control dislocation was a primary concern in designing the DMIs used.

This decision was reinforced by Emerson and Egermann's suggestion that removing this basic confounding factor would allow audiences to more readily access the aesthetic aspects of the performance (illustrated in Figure 3.2). They suggest that gesture-sound causality is 'the foundation for understanding the functioning of the instrument, which then underlies such higher-level evaluative concepts such as perceived liveness and skill' [65].

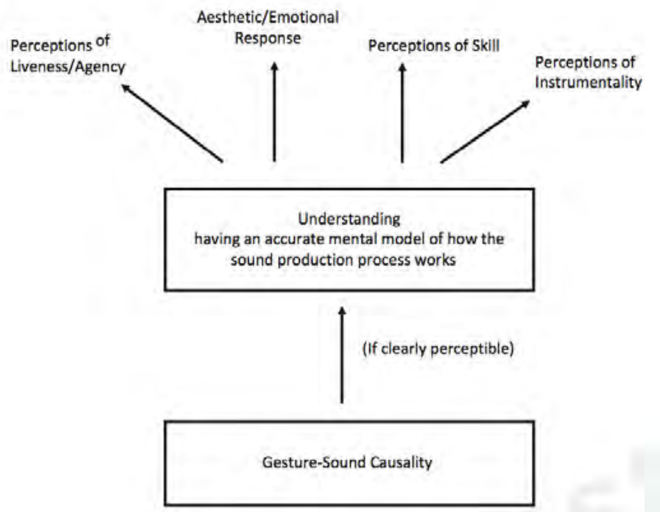
The musicians played music they had composed. In an attempt to make these studies aesthetically interesting for the audience, musicians composed the music played in these studies. This created a performance scenario where the musicians had to execute music of their own creation, and had a greater investment in what they were playing.

Although I invited this creative collaboration, some general instructions were given on the duration of the piece, and parameters of style. For example, the musicians in Study 1 were asked to create 2 pieces, each 3-5 minutes long, and that one should be 'conventional' and the other 'experimental'. (In all instances, the musicians were encouraged to permit these loose parameters according to their personal music practice.)

Though this created a degree of variability between the performances, it did contribute to the ecological nature of this musical environment - the musicians were able to creatively contribute and use their available and learned skill to create a musical experience.

Additionally, each musician was a professional with at least 5 years of performance experience. As such, there was a certain amount of practice, training, performative know-how, artistic rigour, and physical motor skill development that meant these performances would achieve a certain level of quality, and that none could be dismissed out of hand simply because the musician wasn't any good.

Figure 3.2: Emerson and Egermann's model of the DMI reception process [65]



3.4 Understanding audience perception through enjoyment and error

In this research I specifically sought moments of self-reported audience experience of *enjoyment* and *error*. In the course of research I did not instruct or discuss with audiences what 'enjoyment' and 'error' meant; instead, in an effort to gain insight into how audiences define these in the context of a musical performance I instead allowed them to define these for themselves and respond as they saw fit.

Though terms such as 'engagement' are common in the HCI literature and there is intense interest in the nature of emotion within this community, there is far less emphasis on enjoyment. Monk et al. go so far as to term it 'HCI's unbeloved child' [142] due, they say, to the inherent subjectivity of the term. Subjectivity lends an element of fuzziness to user experience, and produces results that could be discounted as too personalised or context-dependent to be useful.

Nevertheless, there is discussion of enjoyment within this context. Brandtzaeg et al. define enjoyment as 'a subjective experience' and further assert that it 'may be understood in relation to theories of motivation' [34]. The first part is useful here, as this body of work engages with the inherent subjectivity of the musical experience. However, the second half of the definition is less useful, as it indicates that the human-computer interaction is what we're most interested in, whereas this context privileges the audience's perception of that interaction.

Indeed, it is the preoccupation with the relationship between the person and the computer that makes much HCI nomenclature less useful here, as it emphasises the performer-instrument relationship and tends to leave the audience to one side. User experience research offers many definitions of 'engagement', but these are generally centred around the human-computer relationship, and offers little in terms of what this might mean for audiences.

Engagement definitions often do have a useful common element: The use of the terms 'interest' and 'attention' [153]. Even in this sense, engagement does not capture the same qualities as enjoyment; one can be attentive and interested in an event while not finding it enjoyable.

Further, these definitions of engagement are largely aligned with positive *valance*, which refers to the emotional response associated with a given stimulus. But, as Latulipe et al. point out, 'it is clearly possible for people to be attentive and interested with negative valence (think of disgust, terror or anger)' [125]. They also state that querying valence steers audience study toward the territory of affective computing, and 'much of this work is aimed at presenting or generating affect' and does not necessarily apply to audiences. [125]

I want stress that, in this thesis, seeking audience indications of 'enjoyment' was not done with the intention of generating any understanding, categorisation or translation of the emotional content of that enjoyment, and does not seek to identify a cause and effect relationship between elements of musical performance and any emotional state. This is not to say that the emotional mechanics of enjoyment are not potentially fascinating, but at the early stages of this kind of audience study, these lie outside the scope of this research.

Within the HCI domain, 'enjoyment' is typically seen as an additional aspect of the user experience; people do not necessarily have enjoyment as a primary goal when using a computer. However, watching music performance and enjoying it is as old as music itself; this is not an emergent feature. Because audiences have an intrinsic cultural understanding of what it is to watch music and experience with enjoying it, it can be assumed that every spectator has a base level of understanding of and experience with 'enjoyment' in a musical context.

The fuzziness and personalised nature of enjoyment is central to the way in which this research is carried out. This thesis departs from the position of musical performance being a complex and multimodal experience, and that the flattening of facets in pursuit of a more straightforward outcome potentially obscures the aspects that are most revealing, and yet undiscovered.

3.4.1 Seeking enjoyment to understand error

As explored in Section 2.2, error is a multifaceted and nuanced concept. Though the Western classical tradition classes 'error' as anything in that performance context that goes against the vernacular and Werktreue of that musical genre and, therefore, sees little value in it, there is suggestion in other areas of musical scholarship for these diversions being the seat of musical style, and a source of creative opportunity.

The most fundamental question, however, is whether error exists *exclusive* from enjoyment — in other words, whether error is perceived when there is no enjoyment. If there is in fact a more nuanced relationship at play, these indications in a performance have the potential to lend insight to performance features that audiences find confusing or boring.

It is precisely because of error's unclear relationship to enjoyment that it is worth seeking and positioning it alongside enjoyment to gain insight into how the two concepts intersect. If a binary nature exists — that is, if error is detrimental to the audience experience — then this would be a useful finding, and shed light not only on what error means in this context, but also enjoyment. Conversely, if no clear binary relationship exists, this has the potential to lend insight to how audiences understand error, and its impact on the musical experience.

In presenting these terms of 'enjoyment' and 'error' to the audience, special care was taken to avoid these being construed as binary terms (this is explored more fully in Chapter 4). I was acutely aware that these terms could be interpreted as opposite ends of a scale, and that it was vitally important to mitigate this interpretation to gain insight about how audiences understand and experience error and enjoyment severally, and how the two terms inform one another.

Additionally, it should be noted that the pursuit of 'error' in this research is not simply an attempt to just get musicians to make mistakes on stage and then see if the audience catches them, or to make interfaces that are simply difficult to play. Instead, understanding how audiences identify of error in the absence of musical vernacular has the potential to indicate where DMI performance can take advantage of the expectations of the audience, and where it might be able to communicate more clearly. Study 2 (Chapter 7) and Study 3 (Chapter 8) examine how the design and function of the DMI's form factor might aid in, add to, or hamper audience understanding, and are not simply props to challenge the motor skills of musicians.

3.5 Data gathering

A major consideration of any body of research is what data is gathered, and how this gathering is carried out. These considerations were a central concern throughout all the studies described by this thesis, and considerations of data gathering techniques and the use of that data form the basis of RQ4[2]. In this section, I examine the two types of data that were gathered: Post-hoc and real-time data. I examine their relative advantages and disadvantages, their current use in related research, and I identify the potential of both types of data to inform one another as well as confounding factors within this relationship. Finally, I then describe how data was collected in these studies in light of the considerations above, and briefly outline how each data set was analysed and used to draw conclusions.

3.5.1 Post-hoc data

In the NIME and DMI communities, the audience is under-studied in general [175]. The audience studies that do exist commonly use post-hoc data (generally questionnaires and, to a lesser extent, interviews) that gather both qualitative and quantitative responses.

The beauty of post-hoc data gathering is that it is a flexible method that can be carried out in a variety of ways, that can be customised depending on the situation, research questions, and the participants. For example, this data collection sometimes take the form of written questionnaires [70] or online feedback forms or surveys [32], but can also take the form of interviews

[2]RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

with audience members [124], or even asking participants to watch video of their interaction and asking them to describe what they were doing, thinking and feeling at the time [25].

These are all methods of polling the opinions of participants, and may include quantitative methods (such as marking numeric scales or rank ordering), as well as qualitative data (in the form of long or short form unstructured narrative answers). Quantitative data can be statistically analysed in order to draw conclusions about general sentiment, but the meaning of these conclusions is often informed by the consideration of qualitative data. Qualitative data is a rich source of contextual and descriptive information (known as *languaged* information). It may be gathered by interviews, questionnaire, conversation (along with a variety of other methods), and is typically analysed using techniques such as thematic analysis [35].

But, the use of post-hoc data in the DMI sphere is remains somewhat under-scrutinised. Perhaps because the focus on the audience is relatively new, there is little critical discourse within the community about how this kind of data collecting should be done, and there is no engagement with questions presented by post-hoc data as a research method.

There is considerable discussion that has existed since the 1970s in the psychology domain about the reliability of the human memory. Loftus et al. notably established that eyewitness recollection, particularly about details of speed, time and duration, are notoriously unreliable [129], and this was countered by Bekerian et al. [19][3]. Nevertheless, this continues to be an active vein of discussion, and though a discussion of the veracity of human memory is beyond the scope of this thesis, decades of scholarly disagreement can reasonably be taken to mean that human memory is, at the very least, inconsistent.

In the study of live audiences in the DMI domain, real-time data is conspicuous by its absence. Though it is entirely possible that the processing of a musical experience and the making of meaning happens after the fact, any discussion about the role and content of real-time data in this community is curiously missing, especially when considering music's time-dependent nature and the demonstrated inconsistency of human memory.

[3]Interestingly, the last author of this paper is John Bowers, who went on to become a DMI practitioner who has been a leader in pushing the limits of the notion of instrument, as cited in the review of contemporary DMI performance practice in Chapter 2.

This is not to suggest that post-hoc data is not valuable. Post-hoc data is an easy and cost-effective collection method that can collect both quantitative and qualitative data. The qualitative aspects are particularly useful, as they provide rich contextual data and descriptive language about the participant's experience. This is also not to suggest that post-hoc data is not 'true', or that 'truer' data exists; certainly our memories and experience develop and change over time, so post-hoc data might mirror this natural decay in our recollection of experience. The question still remains, however, of what real-time data might mean for this type of study, and what other facets of the audience experience they might reveal.

The lack of scrutiny is made more pertinent by the fact that music's time-based nature. The musical experience unfolds over time, and as descriptive and rich as qualitative post-hoc data might be, there is no way to tie this to a particular moment in a musical experience. The question then emerges: What important insight might data post-hoc be missing, and how might this other facet of insight lend dimension to recalled experience? Further, if we consider opinions in-the-moment and after, what we might learn about how musical experiences are processed over time?

3.5.2 Real-time data

Real-time audience response has been measured since 1930s where it was first used to gauge audience response to radio, film, and television [140]. These studies took place in a lab, where spectators indicated their reactions with buttons and knobs on hand-held devices. Since then, real-time data gathering techniques have become more sophisticated and integrated into the performance setting in both the HCI and DMI domains, and now include data sets of physiological data (such as eye tracking [17], facial recognition [114], galvanic skin response [125, 205]), verbal and non-verbal feedback [3, 57], as well as the measurement of crowd behaviours, such as applause [12, 37].

These methods that rely on physiological data have the advantage that they measure bodily phenomena and therefore are not reliant on descriptive language, which risks being misinterpreted, or not an accurate representation of the participant's true experience. Methods to measure crowds, such as applause meters, can only gauge the reaction of a group of people, but have the advantage that they are more resistant to skewing by one overly enthusiastic individuals.

Measuring physiological feedback, however, has major drawbacks for this body of work. Firstly, there is the issue of how it fits in to the live experience and the established performance frame. For example, devices such as head trackers are cumbersome and intrusive, and eye tracking can only reliably be done when watching a video recording. This is a major consideration, as so much of the study design has been oriented towards preserving the multimodality of the concert experience, and additional devices have the potential to distract from and disrupt the performance frame.

A second related issue is that these systems, such as infrared body tracking or facial recognition, often require specific environmental conditions to work properly (such as lighting), or the use of specialist equipment (such as a night vision camera). These add cost and complexity. Additionally, they then orient all design decisions of the performance environment towards satisfying the needs of the data collection method, instead of privileging the construction and maintenance of the performance frame.

Third, and most importantly, there is the issue of the suitability of these methods for the research questions within this work. Passive facial recognition may collect plenty of data within this context, but this does not access the aspect of the conscious judgement of the audience about their experiences of enjoyment and error that are central to these research questions. These methods, such as SHORE [1], are designed to reveal the emotional state of audience members through facial analysis. However, this valence (as described in Section 3.4), though potentially revealing, is not within the scope of this work; rather, I am interested in the more subtle and culturally-dependent aspects of audience judgement, and these currently cannot be inferred by computational means in this context.

A study that is particularly relevant to this research that illustrates some of these drawbacks was one done by Stevens et al. [183] using pARF, a system comprised of 20 hand-held (PDA) computers programmed to gather time-series 'arousal' data. Participants indicated their emotional state with a stylus on a 200px x 200px grid on the device's screen, and their response was measured at a rate of 2Hz. The devices were distributed to 20 individuals in an audience of 200 for feedback during a dance performance.

Though rigorous analysis of the real-time data gathered with pARF was performed, drawbacks to this method are apparent. First, the pARF system supports up to 20 devices (only 18 were used for the study), meaning that a <10% subset of the audience used it, a small sample that is generalised to a much larger crowd. Secondly, no post-hoc data was collected alongside pARF (except for demographic details), which also leaves open the question of the difference in insights this method might have when compared to post-hoc data.

Real-time data provides the moment-to-moment response, but lacks the reflective context that lends insight into the reasons for audience reactions in real time. On the other hand, post-hoc data has rich descriptive data that is entirely contextual, but is entirely reflective, and this reflectivity raises questions about how real-time data might differ, and why. In addition, no features of post-hoc data, except possibly in techniques such as that used by Bilda where participants described their actions as they watched them back on video [25], can be linked to a precise point in real time.

It is clear that post-hoc and real-time data are potentially complementary, and that there is the potential for insight with a combined methodological approach. In response to this need for both descriptive and contextual data, I developed a combined methodology that used post-hoc questionnaire data, as well as real-time audience response via a software system called Metrix that I designed and developed specifically for this purpose.

3.5.3 How data was gathered

Post-hoc data gathering

The post-hoc strategy I adopted was a two-part questionnaire, made up of 'post-performance' surveys, as well as a 'post-concert' survey. The former were short, filled out directly after each performance ended, and asked targeted questions about what the participant had just seen. The latter were longer surveys that collected demographic detail, and asked more comparative, reflective questions about the performances as a whole. (See Appendices C, I and M).

The reasoning that lead to this plurality of questionnaires was as follows:

Capture impressions as soon as possible, as quickly as possible. There were short (5 minute) surveys after each performance to capture immediate impressions. This was to prevent answers being affected by memory decay, and controlled for precedent effect (meaning that it was obvious to the participant what they were responding to, because they had just watched it — precedent effect is, of course, always a consideration, which I discuss later.)

Separate quick impressions from more general reflections. The separation of the 'post-performance' and 'post-concert' surveys allowed for a clear separation of direct reflection on a specific experience, and wider reflection on and comparison between the performances.

Quantize some aspects to find directions of inquiry. Along with targeted qualitative questions, there were also targeted quantitative questions (such as numeric scales and rank ordering) that could be easily analysed in order to provide an indication, on analysis, of impressions the audience had and trends in their opinions. The rank ordering was particularly useful, as it provided a wider view of the performances and additional context by inviting the participants to compare, which is useful in addition to the rankings of enjoyment for individual performances.

3.6 Approach to research and data analysis

In this section, I describe the methodology I adopted to analyse and make meaning of the data collected throughout this research. First, I describe my research goals and theoretical position, as these factors are deeply influential on the outcomes of this work. Next, I describe the methodology I adopted for understanding the qualitative data, and finally, I discuss the handling of the quantitative results.

3.6.1 The research context

The influence of any researcher's existing frame of reference on the resulting analyses is profound, and as such this methodology would be incomplete without a transparent investigation of my goals and values as a researcher in this work, and my theoretical approach.

Research goals

This research was approached primarily as exploratory study. Though several centuries' worth of literature exists that discusses and describes the human experience of art and music, the presence of technology in the artistic experience has been deeply is confounding, and it is still unclear how this has impacted the way that we understand the aesthetic experience. Though inquiry into the influence of computers on our experience of art is an active area of research, it has only existed for a few decades, and as a result the way that technology affects our artistic experience is still largely unclear.

This is compounded by the current deficit of audience study within the DMI research literature. Though existing literature in tangentially related fields certainly informs the context of this work (as described in Chapter 2), too many factors are unknown in the DMI context for this research to result in formalised models or theories that are widely generalisable

For these reasons, this research takes an open and exploratory approach, with the goal not of formalised outcomes — this is unrealistic, given the flux of culture and the newness of this kind of research. Instead, the goal of this research is to test ways of understanding the audience, and gain insight that will indicate fruitful avenues of future study.

Theoretical approach

At the beginning of this research, I again found myself caught between the goals of scientific research (demonstrable, reproducible results), and what I know art and experience to be (perhaps underpinned by commonalities, but meaning and methods constantly in flux).

In order to resolve these two aspects, I adopted a post-structuralist approach to this work. As a researcher and artist post-structuralist writing has been deeply influential on my world view and understanding of audiences. Though interesting, an in-depth description of the application of post-structuralist thinking in this context is outside the scope of this thesis, but it is useful to highlight how this approach has influenced the ways I have undertaken this work, and therefore its outcomes.

Post-structuralism is a late 20th century school of thought originally applied to literary theory, and later extended to wider cultural theory. It can be briefly summarised in this way: Where Structuralism assumes a structure (sometimes referred to as grammar) that underpins cultural experience, post-structuralism 'rejects the idea that meaning is generated and guaranteed by an underlying structure' [184, p. 275]. Instead, post-structuralism assumes *indeterminacy* of meaning; in other words, that meaning is *con-*

stantly in the process of becoming. This is not to say that post-structuralism assumes no meaning, and no truth; rather, the post-structuralist assumption is that 'there is no final arbiter of such decisions' [10], and that there may be a plurality of meanings that may change over time — all equally true.

This approach accepts the mercurial nature of art and experience. I have not approached the audience experience as a single knowable, stable quantity, and do not seek to define it as such. Rather, I am interested in how audiences understand and report their own experience of DMI performance, and examining how these reports converge and diverge among a group of individuals under different experimental conditions in order to gain insight into the lived audience experience as it exists and was reported at the time of observation. As there is currently a lack of discourse within the DMI performance and research communities that tie this art form to the socio-technical aspects of the culture in which it resides, this is a useful approach.

The question then remains of the role of quantitative data in such study. If a post-structuralist approach rejects stable meaning (and, by extension, scientific objectivity), then it is logical to ask why one would even bother collecting numerical data.

Quantitative data, in this instance, was used as a signpost of where meaningful insight could be found — statistical significance in this sense is a useful tool. However, statistical significance among one group of people can do no more than indicate a potentially interesting phenomenon that may hold further insight, that was observable in the time and place in which the study took place. Assuming a potential plurality of meaning does not disregard the possibility of trends emerging that may be useful and insightful. In the synthesis of this data, quantitative results are not used to create larger generalisable statements about audience experience, but rather as clues about how DMI design affects the very murky world of the self-reported individual experience of DMI performance, and provides a metric against which to understand the real-time findings (and vice-versa).

This research is largely centred around reported 'enjoyment' by audience members (quantitatively, qualitatively and in real time), and accepts that there is a multitude of interpretations of that word, and just as many ways that something may be enjoyable. The goal of this work is not to probe the nature and form of that enjoyment in a DMI context, or the form of enjoyment of music in general. Instead, the goal is to observe where and how experimental conditions may have moved that enjoyment needle, and to infer the causes of these effects.

In the real-time data domain, the same may be said about querying for error: That it indicates a stimulus that was interpreted as an error at this time and in this context by this audience, and that examining it may therefore be insightful. Further, it lends dimension and contrast to our understanding of audience 'enjoyment'.

3.6.2 Methods of qualitative data analysis

Qualitative data was collected during all studies via written questionnaire, and analysed using thematic analysis as described by Braun and Clarke [35].

Thematic analysis was chosen specifically because of its flexibility. Many methods of qualitative analysis exist (such as grounded theory, discourse analysis, among others), each with distinctive advantages and drawbacks. Unlike other methods, thematic analysis is not tied to any pre-existing theoretical framework, allowing it to retain flexibility as a method of analysis. Because of the exploratory nature of the studies described in this thesis, I required a qualitative methodology that had inherent flexibility, and thematic analysis has this advantage. As this particular research field lacks precedent of in-depth audience study, the flexibility to observe emergent themes was extremely important, and something that thematic analysis, with its lack of theoretical assumption, could provide.

It is precisely because of its flexibility, however, that thematic analysis requires a thorough description of the approach taken. For instance, I opted in my analysis to focus on producing an analysis that describes the entire data set (as opposed to, for example, analysing the responses of individuals, or a deep exploration of one particular theme). This does result in an analysis that is broader instead of detailed, but in the under-researched area of the audience in the DMI literature this is a useful approach, as the goals of this work are exploratory in nature.

The distinction of inductive vs deductive methods is also an important aspect to highlight. This research used an inductive, or 'bottom-up' approach, in which the themes are strongly linked to the content of the data, and does not assume a pre-existing coding framework. (By contrast, a deductive, or 'top-down', method is driven by the analyst's theoretical interest in a particular area, and looks for patterns related to established themes.)

I coded the data with semantic (expressed verbally) as opposed to latent (underlying meaning) codes. The difference is described by Braun and Clarke as follows:

> *If we imagine our data three-dimensionally as an uneven blob of jelly, the semantic approach would seek to describe the surface of the jelly, its form and meaning, while the latent approach would seek identify the features that gave it that particular form and meaning.* [35]

A semantic approach, therefore, is more appropriate for this analysis, as there is no existing theory. Additionally, a semantic approach allows for a description of the form of the audience experience, and lay the groundwork for more in-depth, latent inquiry in further research.

Thematic analysis was carried out after Braun and Clarke, using six steps:

1. Familiarising yourself with the data
2. Generating initial codes
3. Searching for themes
4. Reviewing themes
5. Defining and naming themes
6. Reporting findings

The results of Steps 3, 4 and 5 are where the subjective judgements and opinions of the researcher are most influential, and maps of these themes are included in Appendices D, J, and N. Each study chapter includes a brief description of the process, and emphasises the final step, reporting findings.

3.6.3 Methods of quantitative data analysis

In each study, participants were asked to rate performances on certain criteria from 1 (least) to 5 (most). For Studies 1 and 2 these were Enjoyment, Interest and Understanding, and for Study 3 this was simply Enjoyment. Additionally, participants were asked in their post-concert surveys to rank order the performances in order of preference.

These quantitative results were analysed using statistical methods that were determined by two factors: Data pairing (whether the ratings or rankings of the two compared sets were made by the same audience member), and data normality (whether the data had a Gaussian distribution). First, a test (such as a t test or one-way ANOVA) was used to determine whether statistically significant differences existed. If differences were found and if appropriate, post-hoc analysis was performed to determine which aspects were significantly different. All analysis was performed using Prism 7[4]. The charts that appear in this thesis were generated as part of this process.

Summaries of the quantitative results are reported in each study chapter.

3.6.4 Method of real-time data handling

Real time data was collected via a web app I developed for the purposes of this study, called Metrix (see Chapter 4 for an in-depth description). Metrix runs on any internet-enabled smart phone, and records audience members registering indications two states via discrete button presses: 'I am enjoying this' and 'There was an error'. This limited response system was designed so audiences would not be distracted by answering multiple questions or keeping track of a complicated interface.

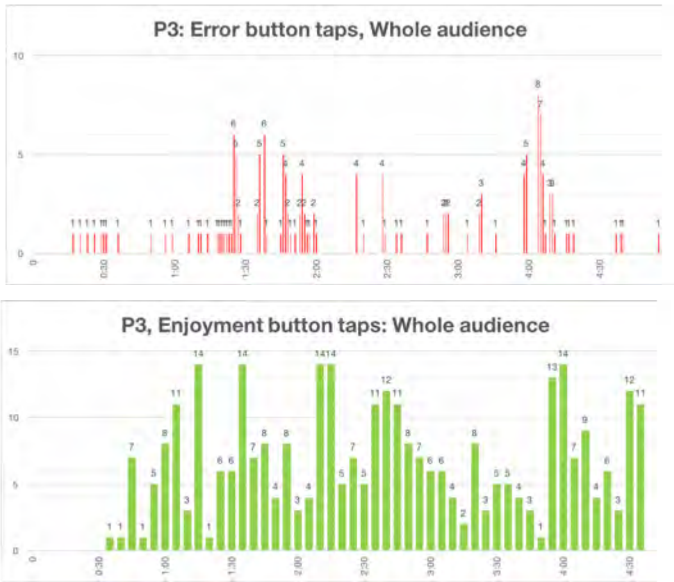
I allowed audiences to select the moment of judgement. If they felt their judgement aligned with 'enjoyment', they were invited to press the button as often and in whatever way they saw fit. If they detected something they classed as an 'error', they were invited to press the other button.

The real-time data set was examined to find trends that appeared at specific points in the data, of both enjoyment and error. By viewing these button presses as histograms I could then detect trends of audience agreement, and also observe patterns within these trends (see Figure 3.3). Then, by syncronising the time series data with the video footage, the video was examined for error at the points where there were times of audience agreement, in order to extract performance features that contributed to audiences reacting with an error or enjoyment tap.

The real-time data collected by Metrix recorded the time stamp, to millisecond resolution, when a participant tapped either of the available buttons. As I was looking for trends of agreement, 1ms proved to be too fine a

[4]https://www.graphpad.com/

Figure 3.3: Sample histogram of real-time 'error' and 'enjoyment' audience data from Study 1.



resolution for video examination. Instead, the button tap data was divided by time bin in order to observe trends of agreement. Because 'error' indications proved to be more rare and sudden than enjoyment (which appeared as a cumulative effect over time), grouping the error taps into 1s bins made the audience reactions more obvious, and indicated parts of the video that could be observed for features. The 'enjoyment' taps were grouped into 5s bins, which revealed the cumulative effect over time (multiple taps within this 5s bin were disregarded to avoid over-zealous button tappers from skewing the results).

3.7 Research questions of each study

The final consideration was in the division of the inquiry of this research over the three studies.

This research is ultimately about understanding the nature and function of error in live performance, informed by how audiences perceive this kind of music given its lack of vernacular and playing tradition, compounded by the confounding nature of the presence of the computer. To truly gain insight into error in this context, some of these confounding factors had to be addressed first.

The inquiry was divided across studies in the following way:

1. Study 1 directly queried the relative effects of musical style and instrument familiarity on audience enjoyment. Currently, as explored in Section 2.3, this confounding factor has not been thoroughly explored, and without some insight into this question this confounding factor would linger throughout this inquiry. As such, the motivation for this study was to disambiguate the influence of the playing style from the influence of the instrument's familiarity, in order to gain clarity about this relationship.

   Still, the use of Metrix for real-time data in this study offered the opportunity to gain insight into specific areas of the research questions: RQ1[5] is directly impacted by this study, as the post-hoc and real-time data analysis provided the first insights to this question. Similarly, the combination of data also supplied useful knowledge on RQ 4[6].

2. Study 2 sought to gain more specific insight into the function of gesture. As explored in Section 2.3, much has been written about gesture in the DMI literature, but there is not much specific information on what communicative gestures should look like, or how some gestures might communicate better than others. This study was motivated specifically by wanting to gain insight into these topics.

   This study also presented an opportunity to gain insight to some of the other research questions. Since there was a laptop performer that was playing entirely free improvisation, the results have some influence on RQ1, and the presence of the combined data gathering methodology also provides insight for RQ4. Additionally, since the instrument was designed with the opportunity for error in mind, this means it was also relevant to RQ2[7]. Further, since this study was a comparative case between two versions of the instrument, it also contributes knowledge to RQ3[8]

[5] RQ1: With no external stylistic frame of reference, how does the audience perceive error in DMI performance, and how does this affect enjoyment?

[6] RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

[7] RQ2: What is the impact of visible risk on audience perception of DMI performance?

[8] RQ3: Can the physical design of a DMI affect the performative outcomes?

3. Study 3 departed from the knowledge gained from Study 1 and Study 2, and was an investigation into how a disfluent behaviour in a DMI might produce effects perceivable by audiences.

   In addition to this, Study 3 provided insight into all research questions except for RQ1, as it was necessary to eliminate the confounding factor of unfamiliar playing style for this study. However, this study lent insight to all remaining questions in this research.

3.8 Summary

This chapter provided a description and rationale for the approach I took to studying the perception of DMI performance by live audiences of DMI performance. The aspects of the performance setting I discussed were the live setting, preserving multimodality, understanding and controlling complexity. I also described how I query error by seeking it alongside enjoyment, and how the contrast between these can reveal the nature of error as perceived by audiences in this context.

I discussed the data that was gathered and how that gathering took place, as well as how the data was analysed to produce results. This includes a discussion of my theoretical approach and research goals, as I acknowledge that these aspects of the researcher impact the results.

Finally, I described how the questions in this research were addressed by the three studies described in this thesis. I briefly described the central motivating question of each study, and detailed how they contributed insight to the aspects of RQ1 and RQ2.

In addition to this methodological chapter detailing the design of these studies and the data gathering techniques, chapters in this thesis more fully explore the tools used for these studies. The following chapter, Chapter 4, details the architecture, design, and use of Metrix, the system I used for real-time audience data collection in a live performance setting. Additionally, Chapter 6 details the instruments built for Studies 2 and 3, and the design decisions that supported those studies' experimental conditions.

Chapter 4

Metrix: A system for real-time audience data gathering

This chapter describes Metrix, a system for collecting real-time audience feedback.

Currently, the audience is under-studied in the DMI research community, and the studies that do exist use largely post-hoc data to understand audience perception. As discussed in Section 3.5, post-hoc data is a useful source of rich quantitative and qualitative data, but this data cannot be reliably tied to any particular temporal moment, and there is considerable evidence that our impressions of happenings in the moment change considerably over time.

For these reasons, I was interested in adding a real-time data dimension to this audience research, as the above evidence suggests that there are new insights to be observed. But, I quickly found that an appropriate tool for this kind of data collection did not already exist. In response, I designed, developed and implemented Metrix for my own research, and it is currently publicly available for use by others[1].

Section 4.1 describes the need for this system. As part of this section, I describe other available tools that I researched, and detail the reasons why they were not fit for purpose.

In Section 4.2 I detail Metrix's technical architecture, how it runs, how

[1]Metrix repository: https://github.com/disastrid/metrix

Figure 4.1: Logo for Metrix.

the data is collected from spectators, and how that data is stored and retrieved. I also briefly describe the data processing techniques used to clean and analyse this data.

The Metrix interface underwent three rounds of design iteration. Section 4.3 discusses the specific design decisions, and this iterative evolution of Metrix that was carried out in response to participant feedback.

One challenge I faced with this research was audience onboarding. Section 4.4 describes the process I developed for this, and includes a discussion of usability principles that I employed in order to make the onboarding process intuitive and easy.

Finally, Section 4.5 outlines areas of observed success in using Metrix, and identifies areas for further development.

4.1 Why Metrix was necessary

When considering how to do real-time data gathering in a performance setting, I first looked to existing solutions. After reviewing a number of options, I was surprised to find each of them to possess some combination of the following drawbacks: Completely unfit for purpose, drastically over-featured, expensive, cumbersome, inflexible, or lacking in control over the format of the collected data.

At the outset, I had the following requirements for a solution that would allow me to collect the data I needed:

- An interface that was unobtrusive and extremely minimal, as to not interfere with the concert setting.
- Extremely easy to use and learn
- A web-based solution, that would ideally run on a mobile phone but would not require participants having to download and install an app
- A remote control interface that would allow me to make the interface active or inactive as necessary during the concert
- A method of assigning an anonymous username to each participant, so I could associate real-time and post-hoc data sets
- Data that was collected in a way that would be easily accessed processed (ie, JSON format)
- A system that could reliably support up to 65 users connected at one time (this was the capacity of the performance venue)
- Open source, or very affordable

Existing real-time data collection platforms cluster around a number of use cases. Common uses are:

- Marketing purposes, used to gauge user reaction to, for example, online advertising (Survicate[2])
- Educational purposes, such as providing a means through which instructors can have a real-time sense of student engagement with teaching (Socrative[3], Verso[4])
- For seminar or conference audiences to respond to a speaker in real time, by providing a means for audiences to, for example, tweet their questions during a presentation, or tweet their comments to the conference screen (Feedbackr[5], Microsoft Pulse[6])

[2] `https://survicate.com/`
[3] `https://www.socrative.com/`
[4] `http://versoapp.com/`
[5] `https://www.feedbackr.io/`
[6] `https://www.engage.ms/pulse/`

None of these use cases applies to this research, and as the above applications are all commercial services, I found that none could be adapted to fit my specific needs.

I also investigated consultancies offering real-time data services using proprietary, hand-held devices (such as Robinson Research[7] and Padgett Communications[8]). These devices, however, were:

- Limited in number
- Not customisable
- Had over-featured interfaces for my purposes
- Since they would be new to participants and were somewhat complicated, they would require considerable on-boarding for audiences to be able to use them
- They are provided and the study administered by the consultancies at considerable expense

I was surprised that there was no open source solution that took advantage of ubiquitous mobile technology. For this research, which took place in central London between 2015 and 2017, I could reasonably assume the vast majority of participants would have a mobile phone with them, and the university's wifi network could support multiple concurrent connections. Further, in the current cultural climate we are all intimately familiar with, and expert at using our own mobile phones, so making a solution for mobile would eliminate the gating factor of an unfamiliar or new device supplied by a third party. This combination of availability, familiarity and existing skill seemed to be an opportunity to place this system on a platform that would be already familiar to the participants.

Additionally, developing my own solution would allow me to design an interface that was precisely fit for purpose, and could be iterated upon quickly and cheaply. I would also have full control over the data, allowing me to add or remove fields as necessary, or change the data collected for each study. As a downloadable app would invite the increased complexity of multiple device compatibility, security, maintenance, and installation, creating this

[7] `http://robinson-research.com`
[8] `http://www.pcipro.com/`

application as a web app would mean that participants could access it by simply navigating to a page in their phone's browser.

For these reasons, building Metrix became not only an interesting challenge, but necessary for this research. In addition, since there is an obvious gap for a system to do this kind of audience research reliably and cheaply, it is also an opportunity to develop a research tool that would help not only me, but other researchers in this space. With free and available tools such as Metrix, this kind of audience research could potentially become easier and more cost-effective, and, by extension, has the potential to become more common. Metrix is currently available for use, freely distributed on Github via a Creative Commons Non-Commercial Sharealike license[9].

4.2 Technical description

Metrix is a web app built using Node.js that runs on a virtual private server (VPS)[10]. Metrix's interface is designed and implemented using HTML, CSS and JavaScript. Metrix communicates with connected devices using web sockets, implemented using the JavaScript library Socket.io[11]. Button tap data is sent from the participant devices to Metrix through asynchronous HTTP POST requests, and Metrix in turn stores that data in a database built with MongoDB[12], a non-relational database that stores data in JSON-like documents.

4.2.1 How participants connect

When the study is introduced, participants can join the study by navigating to a URL in their phone's browser (the URL used for these studies was `http://bit.do/arts2`, as this was easier to remember than the IP address of the server). When ready, they enter the study with a button tap.

This button tap assigns their device connection a username that is a combination of two randomly chosen words. Metrix makes an entry in the

[9] `https://creativecommons.org/licenses/by-nc-sa/2.0/`

[10] A VPS is necessary because the Node process has to run continuously, and this is usually not permitted on shared hosting services. Though a VPS does have a monthly cost of about $15USD per month at the time of writing, universities often provide VPS services to researchers.

[11] `https://socket.io/`

[12] `https://www.mongodb.com/`

database with this username, along with the group to which the participant belongs, if applicable[13]. Any time the user taps a button on the active study interface, this button tap is logged in their database entry in the field corresponding to the tapped button.

4.2.2 Data collection, and syncronisation

For the participant, Metrix has two states: Inactive (between performances, when there is no performance taking place and the screen is paused), and active (during the performance, when the buttons are available to tap). Refer to Figure 4.3 for examples. During the active phase, participants are able to tap buttons. During the inactive phase, the performance is over and participants are given time to fill out their short post-performance surveys, and therefore no button tap data is useful.

During this inactive phase, the screen reminds the participant of their username. The username is key in this process, as it allows the participant's post-hoc and real-time data to be associated at the time of data analysis. Therefore, the username is always kept at the forefront, in case it changes (this can happen if the participant closes their browser and opens it up again), or they have not yet noted it on their survey book.

Metrix moves between these states via a **remote control interface**, operated by the study investigator. (See Figure 4.4.) Here, the investigator can make the interfaces of all connected phones active or inactive by tapping

[13]For Studies 1 and 2 the participants were split into 2 groups and had to indicate their group when logging on to Metrix. There was no group separation for Study 3.

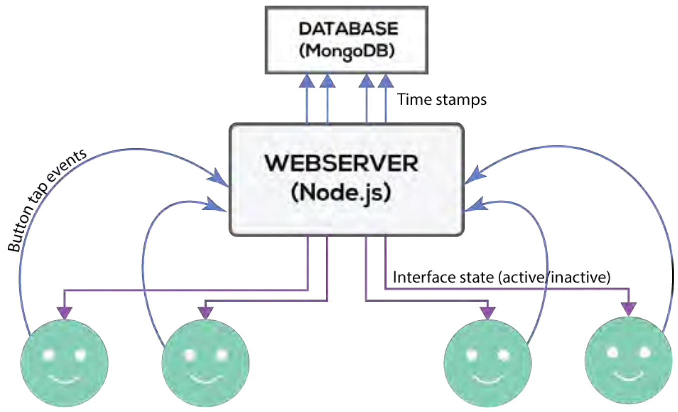


Figure 4.2: Diagram of data flow within Metrix.

Figure 4.3: The live Metrix interface, in two states: Inactive (top) and active (bottom).

a button. These button taps are also logged in the database under a user named ‘remote’, so there is record of when the performances were started and stopped.

The data that Metrix collects is **time stamps**, in epoch time[14]. This time stamp is generated when any participant presses a button.

When any button is pressed, either by a participant or on the remote control, a message is sent to Metrix from the participant’s device with the time stamp of the tap, the button that was tapped, and the participant’s username. In turn, Metrix updates that user’s file in the MongoDB database with the time stamp data in the field corresponding to the tapped button (if the button’s field does not yet exist, it is automatically created).

The logging of remote control buttons is crucial, as it allows the par-

[14]Epoch time, also known as POSIX or Unix time, is the number of milliseconds that has elapsed since the beginning of Coordinated Universal Time, which began on Thursday, January 1, 1970. See the Wikipedia entry for more details: `https://en.wikipedia.org/wiki/Unix_time`

ticipant time series data to be synchronised with the video footage of the performance, allowing for later evaluation.

The recorded performance footage also needs a point of syncronisation with this button tap data. This was achieved with the use of a dog clicker, a device used in dog training that produces a loud and distinctive click. At the same moment of pressing the button on the remote control panel to activate the interface, I also pressed the dog clicker, producing an audible and distinctive click on the audio recordings of the video footage. This creates a point of reference where the video and time-series data could be synchronised later.

4.3 Interface development

The Metrix interface underwent considerable development over the course of this research.

The active Metrix interface is composed of a screen split in half into two buttons, for discrete indications of 'enjoyment' and 'error'. The buttons are logged independently, and can be pressed together if desired.

Originally I considered implementing sliders for audiences to indicate a state on a continuous scale. I reconsidered this approach after critically examining other real-time systems (such as pARF[183] and Mood Conductor [66]) and concluded that this approach is not necessarily conducive to

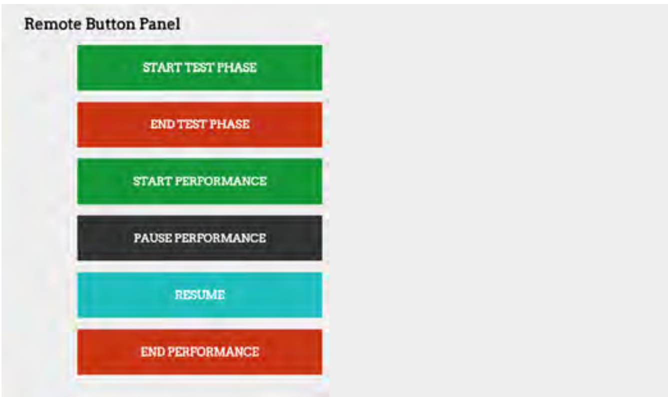


Figure 4.4: The Metrix remote control interface, used to move all user devices between active and inactive modes. The button presses from this interface are also logged on the server, in order to provide a reference of the performance start and stop time.

this particular live context, because a slider requires constant monitoring on the part of the participant using it. Further, if the data shows that a slider was at rest for a prolonged period, it's impossible to tell from the time series data if this resting in a constant position indicates that the participant was consciously indicating that continuous state, or if the slider was in that position for another reason (perhaps the participant simply lost interest in the evaluation, or forgot the interface was there, or didn't understand what the evaluation required).

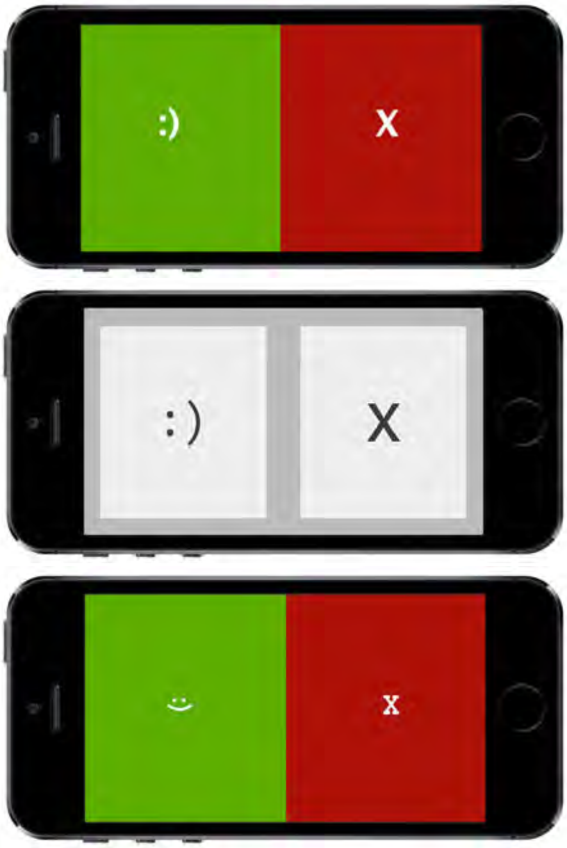
More importantly, a slider indicating a continuum between two states also implies that these states are not only directly opposed, but also to some extent mutually exclusive. As discussed in Section 3.4, avoiding the suggestion of a given binary relationship between 'enjoyment' and 'error' was very important in this context. Further, I did not want to pre-suppose any relationship between the two terms, as this research queries if there is an existing relationship between them, and if so what that relationship is. Though two buttons may also suggest a binary relationship, the two buttons are discrete entities and not tied together at opposite ends of a spectrum, and therefore this binary perception can potentially be mediated through onboarding.

For these reasons, I elected to implement a two-button interface for Metrix. On the left is a button used to indicate 'I am enjoying this'. On the right is a button that is used to indicate 'There was an error'.

Audiences were free to press these buttons as often as they saw fit: They were free to press them frequently, occasionally, or not at all, and to decide when the performance warranted their indication. As the buttons could both be pressed at once, the 'enjoyment' and 'error' states were also independent in function. Additionally, the use of the entire side of a phone screen means that audiences can use two thumbs to tap the buttons, and don't require looking for and finding specific active areas of the screen. Using both sides of the screen does not rely on any button-pressing accuracy, does not require the participant to remember multiple button functions, or and also does not require them to monitor and think about multiple internal emotional states.

The interface has undergone three design phases (see Figure 4.5). The buttons were indicated by symbol instead of text (`:)` for enjoyment and `X` for error) instead of text labels, in order to reduce cognitive load and to act as a quick at-a-glance reminder if the participant had to refer to

Figure 4.5: The three interface iterations of Metrix. From top: Version 1 (2015), 2 (2016), and 3 (2017)



them. The interface was designed to be plain, dull in colour, and generally uninteresting in order to not compete with the performance or impact on the musical experience. Each button darkened slightly on tap in order to provide subtle visual feedback to the participant that their tap had been recorded.

4.3.1 Binary considerations

In Chapter 2 I examined error's negative connotations. I then contrasted these with nuances emerging with the consideration of error from various disciplines, and suggested that this association is worthy of closer examination. This subtlety resulted in my interest in understanding the nature of error, and the first step is determining whether error is associated with a simple lack of enjoyment.

Because I wanted to examine this relationship between 'enjoyment' and 'error', I have been very careful about managing the audience interpretation of the concepts represented by these two buttons, and have taken specific steps to avoid presenting them as opposing terms.

This binary interpretation was mitigated in two ways. Firstly, the two interface buttons function entirely independently, as opposed to, for example, to a slider between the two terms. Secondly, the audience onboarding process stressed the non-binary nature of these two concepts.

Still, some risk remained. Using red and green in the colour scheme, for example, was clearly communicative, but these colours were also complementary; this could possibly reinforce the binary relationship despite best efforts at mitigation.

If the audience was treating these two terms as opposite (such as 'enjoying' and 'not enjoying' instead of 'enjoyment' and 'error'), it would be reasonable to expect that when a participant would tap one button when they were not tapping the other, or that the behaviours around the two buttons would at least have similar features but used at different times. However, this was not the case; the behaviour of 'enjoyment' taps tended to have a normal distribution, gathering to a peak and then tapering off again, whereas 'error' button taps had a very sudden onset and dropoff. Further, 'error' events were not emerging exclusively where there was a lack of 'enjoyment' taps. These trends (which are discussed further in the study chapters) suggest that the audience did not interpret this interface to be presenting two binary concepts, and used the buttons differently.

Still, there was some participant feedback on the interface that suggested that perhaps this colour scheme was confusing this message. To query this, I ran a heuristic design workshop to test this with a group.

4.3.2 Heuristic evaluation workshop

A round of heuristic evaluation was done with a group of 20 people in a workshop setting. Of these 20 participants, 8 identified themselves as having expertise in interface design; 8 indicated they had no expertise in interfaces design; and 4 gave no indication of any kind (heuristic design guidelines recommend a minimum of 4 experts, or 10 novices).

I chose heuristic evaluation because it's a cost-effective and fast way to find usability problems in interfaces [150], and does not require expert

Table 4.1: Average severity ratings, by heuristic, separated by Experts and Non-Experts showing Group Average (Non-Experts includes those who did not indicate expertise)

Heuristic criteria	Experts	Non-Experts	Mean
1: Visibility of System Status	1.63	1.25	1.32
2: Match between system and real world	0.63	0.25	0.53
3: User control and freedom	2.75	2.88	2.58
4: Consistency and standards	1.00	0.38	0.68
5: Error prevention	1.63	0.00	0.79
6: Recognition rather than recall	0.75	0.25	0.47
7: Flexibility and efficiency of use.	1.38	0.63	1.00
8: Aesthetic and minimalist design.	1.25	0.13	0.58
9: Help users recognise/recover from errors	1.50	0.88	1.16
10: Help and documentation	0.88	0.63	0.89

evaluators. I have experience in teaching this evaluation method, and was therefore confident about presenting the concept and process to people of varying levels of expertise who had not done it before.

First I gave a short talk explaining heuristic evaluation (the process, the heuristic criteria, and the severity rating system[15]). Then, I gave the group the same onboarding session given before the study (described in Section 4.4), including screening the Metrix instructional video. Next, I started the app and allowed the evaluators to use it, and assess the interface according to the 10 heuristic criteria. Participants were given a reference sheet of the 10 heuristic criteria and severity ratings, and while assessing the app they were invited to fill out written surveys (the results can be found in Appendix B). The severity ratings for each heuristic are found in Table 4.1.

The heuristic evaluation found no usability problems considered major or catastrophic. However, the written comments and discussion during the workshop did raise some points of interest. I had been concerned that the `:)` and `X` indicators on the buttons were not clear, and were perhaps inappropriate or overly simplistic for the message they were conveying. But, the evaluation group found the communication of these symbols to be clear and appropriate for the message.

The group did feel, however, that the button colours were suggesting a

[15]Metrix heuristic workshop slides: http://bit.ly/2AIN4JJ

binary relationship. It was recommended that I rethink the colour choices.

4.3.3 Interface iterations

In response to the points raised in this heuristic evaluation workshop, I redesigned the interface for Study 2, and implemented a neutral grey interface (see Figure 4.5). The only distinction between the two halves was the symbol used to indicate the button's function in the evaluation.

The feedback from participants of Study 2, however, suggested that, in a performance context, not having a colour distinction between the buttons made the system confusing and more distracting, indicating that colour is a powerful method of quick communication in this scenario. These participants suggested that I implement red and green buttons. As they grey buttons were causing confusion and yet there was no behavioural indication of the red and green buttons indicating a binary relationship, I returned to the red/green buttons for Study 3. Additionally, the `:)` was rotated 90 degrees in order to look more like a smiley face at first glance for clearer communication.

It is important to note that effective onboarding likely affected the interpretation of these buttons, and contributes to the understanding that 'enjoyment' and 'error' are independent concepts. (Onboarding is discussed in the following section.) It was still important to test these assumptions, and to gather feedback both from an evaluation group outside the concert context in order to challenge my own design assumptions.

4.4 Audience Onboarding

Before the study performances, audience onboarding took place.

In the first part I introduced the study and its purpose, and invited the audience to participate. I emphasized that they were free to choose to participate one part (either the questionnaire or the real-time feedback), neither, or both. They then watched the onboarding video[16], and I answered any questions about the system. Lastly, I stated that the buttons did not indicate 'I like it' and 'I don't like it' or 'There was an error' and 'There wasn't an error', but instead were two separate concepts.

[16] Metrix onboarding video: `https://youtu.be/lijwOLO7qOM`

After the video and questions, I activated the system for a test phase. During this phase the screen was active and participants could practice tapping the buttons and seeing what the interface did, but the data wasn't logged. This was so participants could see what the active screen looked like, and get some experience pressing the buttons to prevent presses for the sake of curiosity or novelty. This was a successful strategy, as it seemed to address the novelty factor that might have caused participants to record more reactions to the first performances than the subsequent ones. It also meant that participants were less distracted by the interface the first time it became active.

4.5 Indicators of success and points for improvement

Though an in-depth evaluation of Metrix's effectiveness in an audience setting is beyond the scope of this research, it has proved a valuable tool in this investigation process. This section summarises the features that added value, indicators of success, and evidence that participants found it easy to use, and used it. Additionally, this section details points to be improved in future iterations.

4.5.1 Value points

Flexibility

Because it is built using web technologies, Metrix's design and function are very flexible. Experimenting with the interface, changing the design, and changing the way data is sent to the database in order to meet the needs of this specific research context — something that would be impossible, or at least very difficult with a commercial application — has been possible. Additionally, because the only cost has been my development time, I have not been under pressure to use outside services sparingly, and have therefore had the opportunity to test this system and respond to participant feedback in order to test design assumptions.

Additionally, the flexibility of this system means that it could potentially have multiple uses. I have implemented an aspect of Metrix that allows for the same real-time evaluation of video footage using the YouTube API, and then collect post-hoc responses through an online form. Though the research described in this thesis is specifically and consciously done within a live context, the potential exists for Metrix to be useful in a variety of research applications.

Cost effectiveness

Commercial systems to gather this kind of data are often expensive, or administered by costly consultants. Metrix, however, is open source and built with free and open-source technologies (though it does require a VPS to run because of the Node.js processes, but these are typically widely available and inexpensive).

Stability

The technology used for Metrix — Node.js, HTML/CSS/JS, Sockets.io and MongoDB — easily handled 60+ concurrent connections. No load testing has been done for large audiences, but a system built using these technologies can be expected to handle concurrent connections numbering in the thousands. Therefore, Metrix is at low risk of crashing or encountering problems with a musical audience on the scale of that which is typically studied in DMI research.

4.5.2 Areas for improvement and new features

There are some useful features that I have identified through audience feedback that have not yet been developed, but would increase the usefulness of Metrix as a data collection system.

Persistent usernames

In an effort to use the participants' mobile browsers responsibly, Metrix was built without using any local browser storage (for example, a browser cookie that would log the user in if they navigated away from their browser and then returned to the system). As it is, if participants navigate away or close their browser, they are logged out. They are free to log back in but will be issued a new username. In these studies participants were asked to simply note any usernames on their survey book, but a more seamless way to achieve this would be useful. If using local browser storage, this could be implemented responsibly by ensuring that the token expired after a few hours.

Better data identification

Since implementation I have re-written the functionality of Metrix to not only record the time stamps of the buttons, but to also associate them with the performance number that has been triggered by the remote control interface. This means that the system can potentially produce much cleaner data that will not need such intensive processing.

Re-activating the interface

Currently, if the participant was not on the inactive interface screen when the interface was activated, their browser would not receive the message sent by the remote control and they would not be able to join the performance partway through. I have since addressed this by implementing a function where when a remote button is pushed for the first time, the 'activate' message is fired every 1000ms until made inactive again, so a user joining partway through will be taken to the active screen with little delay.

Documentation of required Node environment

In deploying subsequent versions of this system I have run into compatibility problems, such as discrepancies between the version of Node running on the server and the version of Node for which the current version of Metrix is developed.

Going forward, I would like to explore packaging Metrix as a more easily-deployable system (by deploying, for example, via a Docker image), or at the very least to provide up-to-date version documents clearly outlining the environment requirements and how to ensure that a webserver is properly configured.

4.6 Summary

This chapter provided an overview of the technical architecture, implementation process, and design iterations of Metrix, as well as indicators of success and aspects to be improved. This provides context and a point of reference for the studies described in Chapters 5, 7 and 8, which use data collected by Metrix — both on its own and combined with post-hoc results — to inform their conclusions.

Chapter 5

Study 1: Understanding Familiarity

In this chapter I describe the first study of this thesis, which examined the influence of instrument familiarity and musical style on audience perception of DMI performance.

In Section 5.1 I detail the context, motivations, and research questions driving this study. In this section I also explain how these aspects intersect with and inform the larger research questions of this thesis.

Following this, in Section 5.2 I describe the design of the study, and the method I used in carrying it out. This includes a description of the recruitment criteria for the musicians, the study setup, and the procedure followed for data collection.

The quantitative and qualitative post-hoc results are detailed in Sections 5.4 and 5.3, and the real-time results are reported in Section 5.5. These sections include details of the thematic analysis of the qualitative data and the process of video coding.

Finally, in Section 5.6 I discuss the implications of these results and how they inform the central question of this study. I also describe how the outcomes of this study inform aspects of the larger questions guiding this research, and how these insights fed into Study 2.

5.1 Research questions and motivation

This study investigates the factors shaping the audience experience of DMI performances, comparing the relative effects of familiarity with the technical aspects of a novel instrument and the musical style of the performance.

Visual factors play a central role in how a spectator perceives a performance [188, 175, 62, 157, 100, 126] for both traditional and digital instruments (see [87] for a detailed discussion of instrumental interaction from the spectator perspective). DMI performances are often criticised for being visually opaque [175, 62]. Fels et al. [68] proposed the principle of *transparency*, suggesting that the instrument design should allow the audience to understand the performer-instrument interaction. Since DMIs need not follow traditional instrumental modes of interaction [87], considerable effort has been spent on DMIs which deliberately seek to expose the interaction to the audience. Recent work has proposed physical metaphors [51] and visualisations of control processes [24, 160], and audience experience is increasingly a part of the DMI evaluation process [154].

Fyans et al. [71] conducted a study of audience perception of error in DMI performance under different information conditions. Amongst the findings was that, with regards to the Tilt-Synth (an unfamiliar DMI), explaining the instrument before the performance improved the accuracy of the spectators' mental models of the instrument (though it had no significant effect on understanding the performer's intention in playing the instrument). Several participants suggested 'that they enjoyed the performances more because the performer explained the the instrument first. They commented that it helped them understand the interaction and performance.'

This suggestion has yet to be confirmed; no study so far has measured whether understanding how a DMI works improves audience enjoyment of a performance. The pre-concert talk or demo is a staple of many DMI performances, and it seems plausible that greater familiarity with the operation of the instrument might help a spectator relate to the actions of the performer and thereby facilitate greater enjoyment of the performance. However, the influence of instrument familiarity and the influence of playing style on audience enjoyment has yet to be investigated. This study serves to disambiguate these relative effects, and provide insight into this long-standing question.

5.1.1 Related work

Any investigation of familiarity and audience experience must confront a significant confounding factor: musical style. DMIs can be found all along the artistic spectrum, from traditional instrumental models to interactive compositions. At one end of the spectrum, the DMI is often inseparable from the musical idiom and even the specific piece [62, 145] (see also Jorda's discussion of macro-diversity [107] as a measure of stylistic flexibility).

This close bond between technology and musical ideas may be inherent in the design of some instruments, though in other cases it may relate more to the fact that the instrument's designer is its primary (or only) performer. Musical history is replete with cases where instruments developed for one community found distinctive use in another (e.g. saxophone, bandoneon, electric guitar, Hammond organ). On DMIs, diversity of style is also an emergent property of even the most reductive designs when given to many different players [84, 214].

In any case, a note of caution is warranted in audience studies. Given a DMI which is tightly linked to a musical context, effects on the audience which appear to be due to technical design may instead be effects the style of the performance (or vice-versa). In this study, in addition to examining technical familiarity, the effect of musical style is considered along a spectrum, ranging from the extremes of 'experimental' to 'conventional'.

The aesthetic origins of experimental DMI performance was explored in depth in Section 2.1, but it is worth highlighting the parallel experimental and vernacular streams of digital music because the lack of visibility of performance gesture affects *both* sets of genres. Ableton Live performers can encounter as much criticism as experimental DMI creators for visually disengaging performances; live generative visuals are also found across many electronic genres. But NIME and EDM performances engage the audience in different ways and invite different modes of listening, and the popularity of live EDM performances with or without visual accompaniment suggests that instrumental transparency is not a strict prerequisite to an enjoyable performance.

The prior listening experience of the audience is another consideration. Audiences outside the NIME community are less likely to be familiar with electroacoustic improvisation. Even within NIME, most practitioners wear many hats: composer, performer, instrument designer, audience member. In many cases, the musical genres that a NIME community member listens to in their leisure time may only partly overlap with the genres they participate in in their professional practice.

Understanding and instrumental *transparency* may have cultural as well as technical dimensions. The musical experience of the audience may affect their understanding of a performance whether or not they have encountered a particular DMI before. In the next section, we describe a study aimed at disentangling some of these effects, with the goal of providing design advice for future DMI creators.

5.1.2 Research questions

This study confronts precisely this question of the relative importance of familiarity. The central question for this study is:

> What is the influence on audience enjoyment of familiarity with the technical aspects of the instrument, and musical style?

Along with disentangling the effects of unfamiliar instruments and unfamiliar playing style on audience perception, this study provided important insight into the larger questions in this thesis:

- RQ1[1]: This study contributes to this question directly, as it compares audience's reactions to familiar and unfamiliar playing styles by the same musician.
- RQ4[2]: This study contains the first use of the combined data methodology. It directly contributes to this question, as it is the beginning of explorations of how to combine these two data types.

5.2 Study Design and Method

This section describes the decisions made in the design of this study: The venue, the concert structure, the strategy for dividing the audience, and the procedure followed.

[1]RQ1: With no external stylistic frame of reference, how does the audience perceive error in DMI performance, and how does this affect enjoyment?

[2]RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

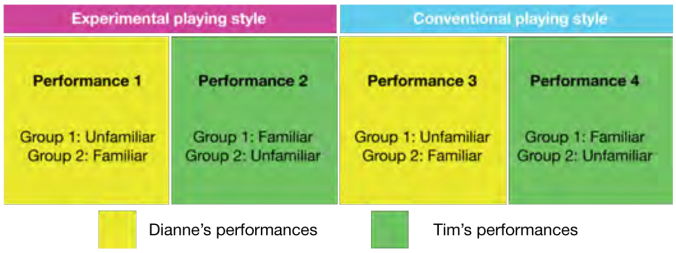


Figure 5.1: Diagram illustrating how the performances provided comparison on two axes: Instrument familiarity, and musical style.

5.2.1 The musicians

It was initially difficult to find musicians who could participate, because of the required criteria for this study:

- They play a novel electronic instrument;
- They have a degree of virtuosity with that instrument;
- They are able to play in both a conventional and experimental style.

When recruiting began it quickly became apparent that the 'conventional style' requirement was going to be a gating factor. There are plenty of musicians who fulfil the first two criteria, but generally these musicians only played in the experimental, 'free improvisation' style characteristic of this kind of DMI performance.

I did eventually locate two artists who were willing to participate: Dianne Verdonk on La Diantenne [199] and Tim Exile on the Flow Machine[3] (see Figure 5.3). Both performers had significant live performance experience with their instrument, and both had previously played their instruments in contrasting musical styles.

To create contrast in musical style, each musician prepared two performances each around 5 minutes in length, one experimental and one stylistically conventional (or vernacular). The performers were free to interpret 'experimental' and 'conventional' in the context of their own musical practice.

[3] http://timexile.com/technology/

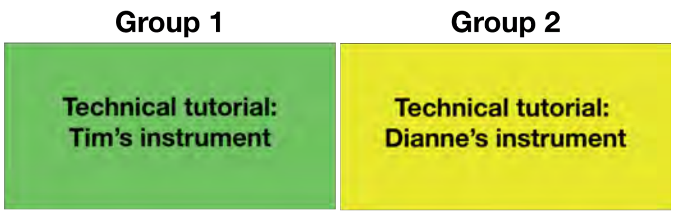


Figure 5.2: Diagram of how the audience was divided

5.2.2 Venue and advertising

This study took place on November 4 2015, as part of an evening concert. The venue was the Film and Drama Studio in the ArtsTwo building, a performance studio with seating for an audience of 65.

The concert was composed of an opening act (a performance on augmented violin), the study act, and a closing act (performances on a hackable musical instrument). The theme of the evening was Innovative Interfaces and was advertised through email lists and social media channels[4]. The audience was advised in these communications that a study would be taking place, and if they wished to participate to bring their mobile phone.

5.2.3 Audience groups and pre-concert tutorials

The audience was divided into two groups (see Figure 5.2). The groups were determined by the colour on the survey books each member randomly received upon entering.

Before the study act of the concert, I presented the study to the audience and invited them to participate. Then, each audience group saw one of two 10-minute instrument tutorials that explained the technical aspects of the instrument and how it creates sound. Group 1 received a tutorial on the Flow Machine, and Group 2 recevied a tutorial on La Diantenne. The tutorials were presented by a member of the research lab to address any bias from meeting the performer prior to the concert. This provided an axis of familiarity: For each performance, half the audience would be familiar with the instrument, and the other half would be unfamiliar.

Each group received Metrix onboarding in a separate room while the other group received their tutorial in the performance space. The onboarding included a short explanation of the interface, the onboarding video, and

[4] `https://www.facebook.com/events/906185819466663/`

the opportunity for questions.

5.2.4 Concert order

No single concert order can entirely address order bias effects, but for consistency we chose to place the experimental performances together at the beginning, followed by the conventional performances. Both instruments were amplified, but since Dianne's was designed to be quieter than Tim's, Dianne performed first. The order is diagrammed in Figure 5.1.

5.2.5 Data collection

The data for this study was collected in two ways: Through 5 written questionnaires (available in Appendix C), and by real-time audience feedback via Metrix (described in Chapter 4).

4 of the 5 questionnaires were short surveys that were filled out immediately after each performance (3-4 minutes were given for this), and asked the participants to reflect on what they had just heard. The participant was asked to rate their Enjoyment, Interest and Understanding of the performance they had just seen on a scale of 1 (least) to 5 (most).

There were also three qualitative questions with space provided for about two sentences. These asked what the participant liked, didn't like, and how they might describe the performance to a friend.

After the final performance and questionnaire, participants were asked to fill out a longer post-concert survey. This asked the participants to reflect

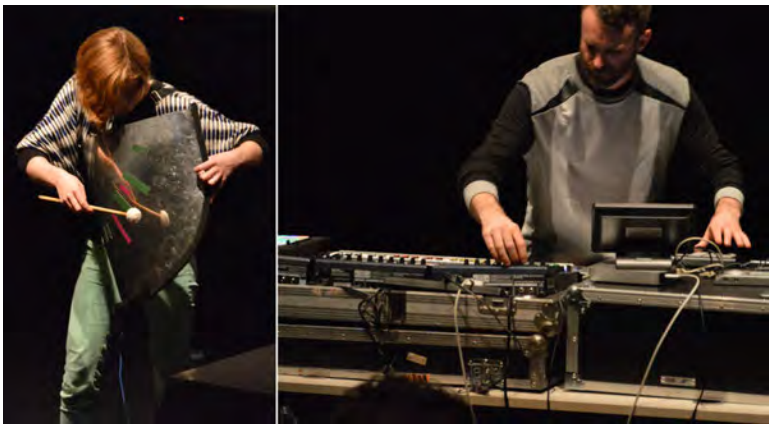
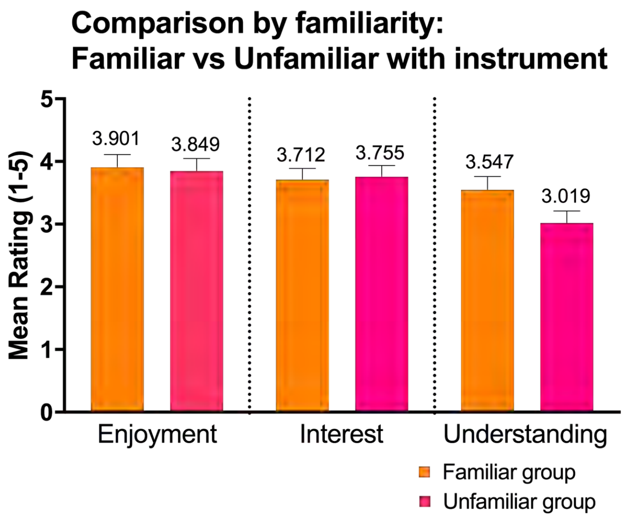


Figure 5.3: The performers. Left: Dianne Verdonk. Right: Tim Exile.

Figure 5.4: Mean Enjoyment, Interest and Understanding ratings across all performances, compared by familiarity.



on the performances as a whole; to rank the 4 performances in order of preference from 1 (favourite) to 4 (least favourite); to rate how well they understood each performer's instrument, and then to rate if they would be able to play it.

This survey also collected considerable demographic detail. This data allowed us to further subdivide the audience, as discussed in Section 5.6.

5.3 Post-hoc results: Quantitative

The audience questionnaire is available in Appendix C. Two aspects of the survey were analysed quantitatively: First, the ratings of Enjoyment, Interest and Understanding from 1 (not at all) to 5 (very much) in each post-performance survey, and second, the rank ordering of the performances from 1 (favourite) to 4 (least favourite) in the post-concert survey.

5.3.1 Comparing ratings by instrument familiarity

First, the ratings of Enjoyment, Interest and Understanding across all performances were compared to see if there was any statistically significant difference between these ratings by those familiar with the instrument, and those unfamiliar with the instrument (illustrated in Figure 5.4).

A Kruskal-Wallis test was conducted to assess this. This test was chosen because the data was nonparametric and the ratings were unpaired. Ratings were statistically significantly different between those familiar with the instrument vs those who were not ($\chi^2(5) = 51.29$, p<.0001).

Subsequently, pairwise comparisons were performed using Dunn's procedure with Dunn's correction for multiple comparisons. This post hoc analysis revealed statistically significant differences in ratings of Understanding between those familiar with the instrument vs those unfamiliar with the instrument ($\chi^2(7)= 93.18$, p = .0004). There were no statistically significant differences between the ratings of Enjoyment or Interest.

To query further, the ratings for Enjoyment, Interest and Understanding by the familiar and unfamiliar groups compared for each of the four performances (visualised in Figure 5.5). A Kruskal-Walis test was conducted on each set of performance data to determine if there were differences in the ratings of Enjoyment, Interest and Understanding between the groups familiar and unfamiliar with the instrument used in that performance.

Post hoc comparison of the ratings of the 4 performances, using Sidak's multiple comparisons test, found that 2 performances had statistically significant differences in ratings between the familiar and unfamiliar groups. These were Performance 1 ($\chi^2(5)$=23.37, p=.0003) and Performance 3 ($\chi^2(5)$=17.97, p=.003).

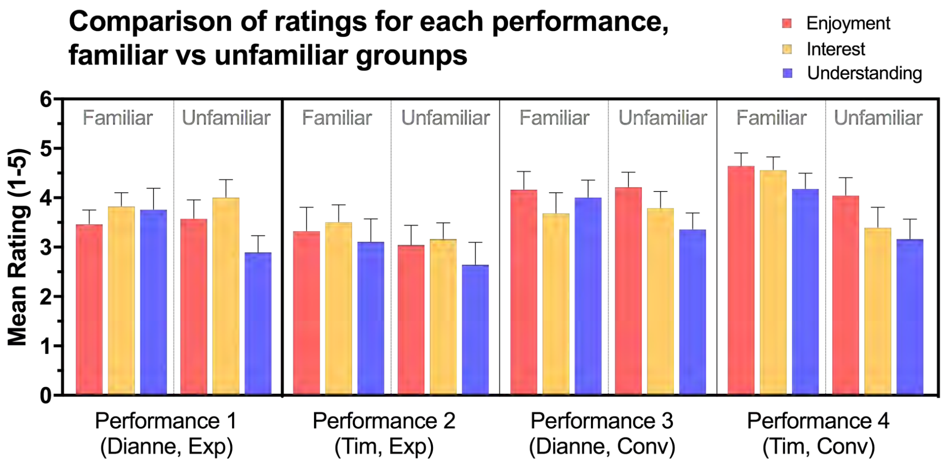


Figure 5.5: All mean ratings for Enjoyment, Interest and Understanding across all 4 performances, divided by familiarity. NOTE: Error bars indicate 90% CI.

Subsequently, post hoc comparisons were performed using Dunn's procedure and Dunn's correction and these post hoc analyses revealed statistically significant differences. For Performance 1, the ratings for Understanding between the familiar and unfamiliar participants were statistically significantly different (familiar mean=3.76, unfamiliar mean=2.893, p=.0018). For Performance 3, the ratings for Understanding between the familiar and unfamiliar participants were also statistically significantly different (familiar mean=4.00, unfamiliar mean=3.357, p=0.0301).

5.3.2 Comparing ratings by musical style

Next, the ratings of Enjoyment, Interest and Understanding for the performances in Experimental (Performances 1 and 2) and Conventional (Peformances 3 and 4) musical styles were compared.

A Friedman test (nonparemetric, unpaired) was used to determine whether there were statistically significant differences between the overall rankings for Enjoyment, Interest and Understanding between the Experimental and Conventional performances.

The test found statistically significant differences ($\chi^2(2) = 59.06$, p<.0001). Post hoc analysis revealed statistically significant differences in the ratings

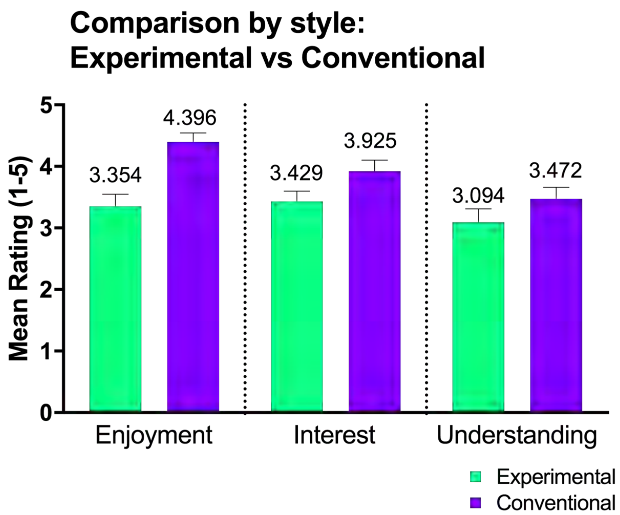
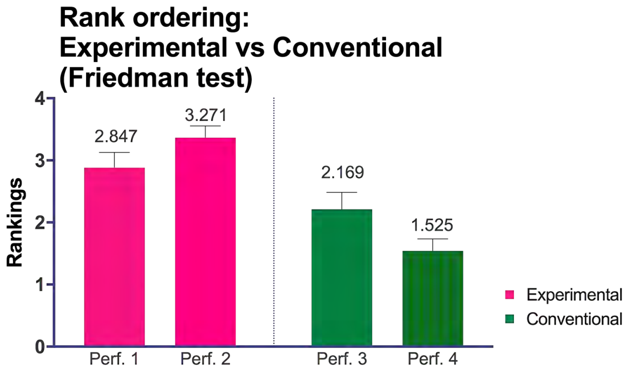


Figure 5.6: Mean ratings of Enjoyment, Interest and Understanding, compared between Experimental vs Conventional styles. Error bars indicate 95% CI.

Figure 5.7: Rank order analysis of Performances 1 - 4, comparing by musical style. Bar values indicate mean ranking. Error bars indicate 95% CI.



between the Experimental and Conventional performances for Enjoyment ($p<.0001$), Interest ($p=.0014$), and Understanding ($p=.0246$). These are visualised in Figure 5.6.

5.3.3 Rank ordering

In the post-concert survey, audience members were asked to rank the performances in order of preference, from 1 (favourite) to 4 (least favourite). This rank ordering data was analysed to determine if there were statistically significant differences when compared by instrument familiarity, and by playing style. I will describe each of these comparisons in turn.

A Kruskal-Wallis test was conducted to determine if there were differences in rank ordering of any of the performances between those familiar with the instrument, and those unfamiliar. Median ranking scores were statistically significantly different between the two familiarity groups ($\chi^2(7) = 75.11$, $p<.0001$).

Subsequently, post hoc comparisons were performed using Dunn's procedure with Dunn's correction for multiple comparisons. Though statistically significant differences were found over the entire data set in this post-hoc analysis, there were no differences of any significance found between the rank ordering of the familiar and unfamiliar audiences for any single performance. The mean values are illustrated in Figure 5.8.

Next, a Friedman test (nonparametric, unpaired) was performed to compare the rankings given to the experimental and conventional performances. Statistically significant differences were found ($\chi^2(3)=59.06$, $p<.0001$). Multiple comparisons were then performed using Dunn's method and Dunn's correction, and this post hoc analysis revealed statistically significant differences between the rank ordering of Performance 1 and 4, Performance 2 and 3, Performance 2 and 4, and Performance 3 and 4 (the results are summarised in Table 5.1 for brevity, and visualised in Figure 5.7).

5.3.4 Comparing ratings of Understanding by instrument familiarity

The audience was asked to rate each performance for Understanding in the post-performance survey, and analyses of this data are discussed above. Ratings of Understanding between the familiar and unfamiliar groups were found to be significantly statistically different.

Two questions in the post-concert survey again asked the audience participants to rate their understanding of each instrument. For each performer, audience participants were asked to rate their agreement with the following statements from 1 (not at all) to 5 (very much):

1. I understood the instrument.
2. I could play the instrument.

These results were divided by familiarity with each instrument and compared with the post-concert ratings of Understanding, to see if there were significant differences between the two groups. Because the data was not

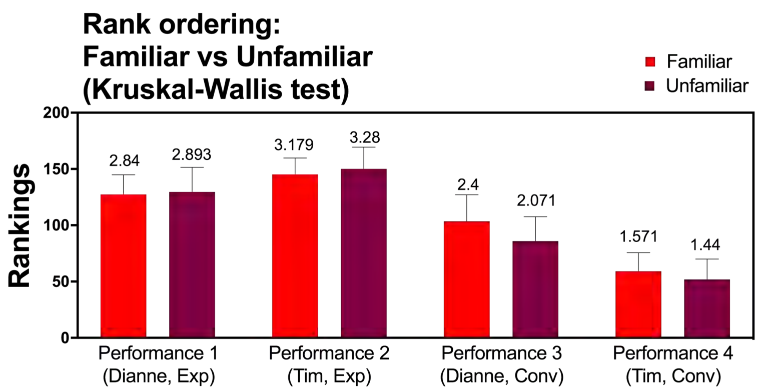
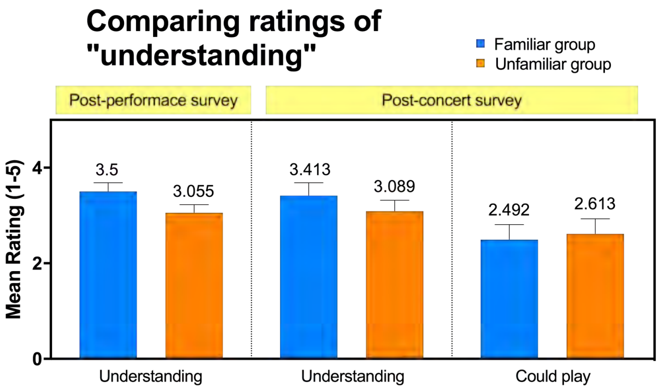


Figure 5.8: Rank order analysis of Performances 1 - 4, comparing by familiarity. Bar values indicate mean ranking. Error bars indicate 95% CI.

Comparison (Dunn's)	Rank sum diff	Adj. p value	Sig?
Perf. 1 vs 2	-22	0.5873	No
Perf. 1 vs 3	33	0.0782	No
Perf. 1 vs 4	73	<0.0001	Yes
Perf. 2 vs 3	55	0.0002	Yes
Perf. 2 vs 4	95	<0.0001	Yes
Perf. 3 vs 4	40	0.0157	Yes

Table 5.1: Summary of post-hoc analysis of rank ordering, comparing playing styles. Statistically significant differences found in 5 of 6 comparisons.

Figure 5.9: Comparing ratings of understanding by familiarity between post-performance and post-concert surveys. Bar values indicate mean ratings. Error bars indicate 95% CI.



normally distributed according to a Shapiro-Wilk normality test, a Kruskal-Walis test was used. (See Figure 5.9 for an illustration of mean ratings).

The post-performance mean ratings of Understanding between the familiar and unfamiliar groups were significantly statistically different for the post-performance survey (p=.001). No differences of statistical significance were found for the post-concert ratings of Understanding between the familiar and unfamiliar groups. Further, when comparing the ratings for 'I could play the instrument' between the familiar and unfamiliar groups, no statistical significance was found.

5.4 Post-hoc results: Qualitative

There was considerable qualitative data collected. In each post-performance survey three questions were asked:

1. What did you like about the performance?
2. What did you dislike about the performance?
3. How would you describe the performance to a friend?

The data corpus assembled for analysis is composed of the answers to questions 1 and 2, for each of the 4 performances.

A thematic analysis was performed on the qualitative data. The process of qualitative analysis described here was followed for all three studies.

I performed this analysis on my own, without the input of other researchers. The reason for this is for clarity of bias and transparency of process. The outcome of thematic analysis is very much reliant on the viewpoint of the analyst, and as a result I want to present an analysis where the biases can be reasonably traced to one researcher. Isolating bias is difficult in a collated analysis that represents the input of multiple researchers. Additionally, as I am the primary investigator in these studies and there is no other researcher adequately invested and familiar with the subtlety of these questions, I was uniquely positioned to do the most in-depth analysis and separate out the most relevant themes.

The process followed was that specified by Braun and Clarke [35]. After familiarising myself with the data I then went through the questions associated with each performance, coding the responses with short words or phrases. There were 58 of these themes in total.

After coding the data, I then grouped the codes into themes using a diagram. This diagram of the themes and the codes they contain can be found in Appendix D. These themes were assembled by grouping codes together that made sense — for example, words related to sound were grouped together (timbre, texture, 'noise', and so on). I then merged associated groups until no more associations could be made.

The result was four themes:

Sound: Descriptors of musical or sonic output (Quote: '*The pitch bends were a bit jarring sometimes*')

Table 5.2: Number of statements on each theme, divided by like/dislike, across all 4 performances for familiar and unfamiliar audiences. NOTE: All values are percentages.

		Sound		Performance		Instrument		Experience		Novelty
		+	-	+	-	+	-	+	-	
P1 (D, Exp)	Fam	40.0	70.0	43.3	10.0	33.3	0.0	3.3	10.0	30.0
	Unfam	37.5	56.3	43.8	12.5	40.6	6.3	40.6	25.0	25.5
P2 (T, Exp)	Fam	62.5	75.0	21.9	3.1	15.6	3.1	21.9	12.5	6.3
	Unfam	60.0	80.0	20.0	13.3	3.3	3.3	13.3	6.7	16.7
P3 (D, Conv)	Fam	53.3	23.3	63.3	20.0	13.3	40.0	40.0	40.0	0
	Unfam	56.3	37.5	37.5	28.1	12.5	6.3	40.6	28.1	12.5
P4 (T, Conv)	Fam	59.4	28.1	37.5	3.1	3.1	3.1	56.3	21.9	0
	Unfam	86.7	40.0	33.3	10.0	0.0	0.0	40.0	26.7	6.7

Way of playing: Comments related to aspects of performance (Quote: '*Interesting to watch him work*')

Instrument: Statements related directly to the interface (Quote: '*Nice looking instrument*')

Audience experience: This theme included statements that were more general and related to the experience of the three aspects of above, such as value judgements and emotional response (Quote: '*Great flow of sounds, I felt like dancing*')

In the data corpus, the majority (97%) of responses could be grouped under these themes. The remaining 3% were responses such as *Not much to be honest.*

The way that the codes of each theme were spread across the data corpus is visualised in colour-coded charts in Appendix D.

The content and frequency of theme statements were analysed, but the frequency of occurrence — expressed in percentage of the audience that mentioned it — was particularly insightful (summarised in Table 5.2). The notable indicators were as follows:

5.4.1 Novelty

Statements related to novelty ('*I've never heard anything like this*') were present in both groups for Performance 1 (mentioned by 30% of the familiar group and 25.5% of the unfamiliar group). For the remaining performances the rate of novelty dropped sharply (For familiar and unfamiliar groups: P2, 6.3% and 16.7%; P3, 0 and 12.5%; P4, 0 and 6.7%).

5.4.2 Sound statements

Sound was the most commented-upon aspect across all performances. Comments related to sound appeared more often as dislikes for the experimental performances, and more often as likes for the conventional performances (see Table 5.2).

5.4.3 Instrument comments

The instrument was the least commented-upon aspect of all performances with the exception of Performance 1.

5.4.4 Experience statements

Across both familiar and unfamiliar audiences, experience statements were applied more to the conventional performances (P3 and P4) than the experimental performances (P1 and P2). However, one half of the total audience (Group 2, familiar with Dianne's instrument) was very consistent in their mentioning of experience in both like and dislike responses, with the exception of P2 for which this group mentioned experience far less (see Table 5.2 for details).

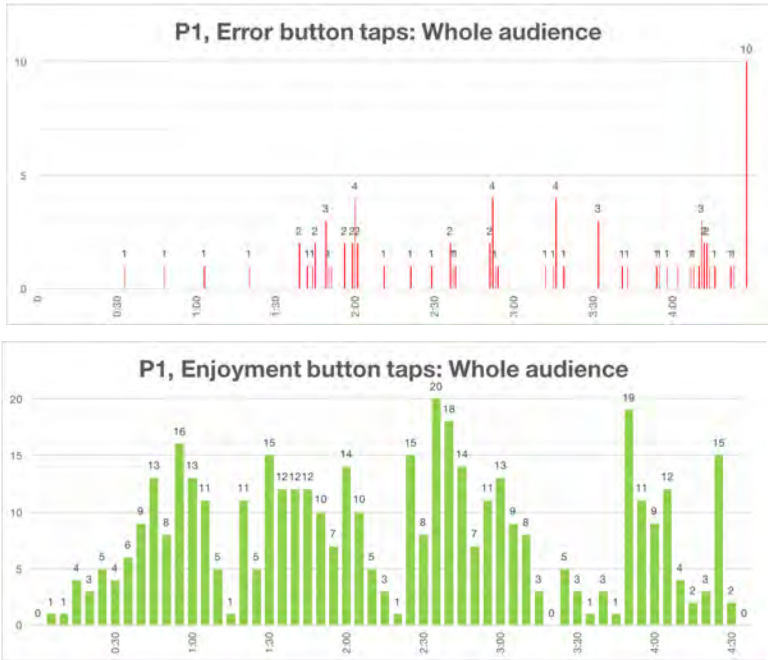
5.5 Results: Real-time

5.5.1 Treatment of the real-time data

The button hits by audience members were time stamped with Unix epoch time, with millisecond resolution. However, the time stamps of 4 participants were rounded to the closest second (the reasons for these are not yet identified, though I suspect this is related to legacy mobile browsers). To address this inconsistency, all time stamps were rounded to 1s.

1s resolution timestamps were found to be precise enough for the 'error' button taps, but still too granular to make sense of the 'enjoyment' data. The 'enjoyment' time bins were widened to 5s, at which point the patterns in

Figure 5.10: Visualisation of 'error' vs 'enjoyment' taps over the whole audience, for Performance 1.



the data became much more apparent. All real-time histogram visualisations of real-time data are available in Appendix E.

5.5.2 Observations of participant button use

There was some concern before this study that audience members wouldn't use the interface, or that use would drop off significantly partway through.

This did not prove to be the case, and participants were enthusiastic button users throughout all four performances (see Table 5.3 for totals of button hits throughout the performances).

Additionally, participants tended to be consistent in their use of buttons; for instance, if an individual was a liberal button-presser, this pattern held throughout their use. Conversely, if their use was sparing, this behaviour was consistent as well.

5.5.3 Observations of 'enjoyment' and 'error'

The first observation was that the buttons were used very differently.

As illustrated in 5.10, the most obvious difference was that the error button was used quite sparingly, while the enjoyment button was used quite liberally. This is also demonstrated in Table 5.3, which shows the button tap totals over the four performances.

The second observation was in relation to the occurrence of events. I use the term 'event' to refer to places in the real-time data where there are spikes in audience agreement. In the 'enjoyment' data, these events tended to be cumulative, rising to peaks over time, whereas in the error data — which, with its 1s time bin, allowed for a much smaller time window for audiences to agree — the peaks tended to appear suddenly, and then drop off.

The third observation of the 'enjoyment' vs 'error' data is that they are not opposing, meaning that error does not only occur where there is a dip in enjoyment. Intriguingly, there are times where 'error' and 'enjoyment' events occur together (in Figure 5.10, for example, there is a spike in 'error' and also in 'enjoyment' taps around the 2:00 mark).

5.5.4 Comparing rates of button taps by familiarity

Before any comparison of button tap rates was done, all tap rates were normalised to taps per minute to account for the varying length of the performances.

In comparing the histograms between the familiar and unfamiliar groups, there were no obvious differences for the real-time data for either the error or the enjoyment taps. To confirm this, the data was compared.

A Kruskal-Wallis test was used to see if there was any difference between the mean number of taps made between the familiar and unfamiliar groups in each performance (see Fig 5.11). No statistically significant differences were found.

	P1		P2		P3		P4	
	Err	Enj	Err	Enj	Err	Enj	Err	Enj
ALL	89	505	190	410	163	397	238	864
Familiar	51	263	88	187	99	228	100	388
Unfamiliar	40	221	102	223	77	174	138	476

Table 5.3: Number of button taps per performance for error and enjoyment.

Figure 5.11: Kruskal-Wallis test for mean rates of error and enjoyment taps, separated by familiarity. Error bars indicate 95% CI.

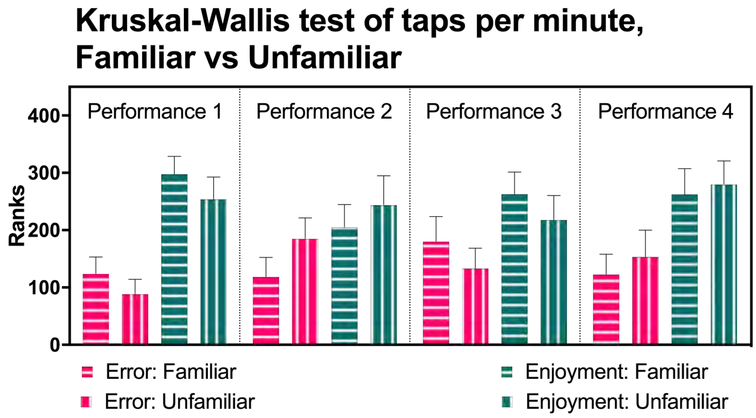


Table 5.4: Chart of Spearman (nonparametric) correlation of button taps and Enjoyment ratings for each performance, separated by familiarity. Significant p values highlighted.

		P1		P2		P3		P4	
		Error taps	Enjoyment taps	Error taps	Enjoyment taps	Error taps	Enjoyment taps	Error taps	Enjoyment taps
ALL	r_s	-0.184	0.420	-0.123	0.580	-0.105	0.252	-0.214	0.239
	p-value	0.202	0.002	0.390	<0.0001	0.491	0.095	0.149	0.106
FAMILIAR	r_s	-0.185	0.483	0.087	0.617	-0.231	0.295	0.086	0.161
	p-value	0.398	0.020	0.655	0.0003	0.302	0.182	0.691	0.454
UNFAMILIAR	r_s	-0.088	0.457	-0.336	0.676	0.011	0.197	-0.463	0.286
	p-value	0.662	0.017	0.126	0.001	0.959	0.369	0.260	0.187

5.5.5 Correlation of button taps and ratings of Enjoyment

A Spearman (nonparametric) correlation was made between the number of button taps made by an audience participant, and their subsequent post-hoc ratings of Enjoyment. The intuition was that if real-time and post-hoc data were related, and if error and enjoyment had a binary relationship, then there would be a negative correlation between the number of 'error' taps and ratings of enjoyment, and a positive correlation between ratings of enjoyment and the number of 'enjoyment' button taps.

Table 5.4 summarises these results. For this correlation, only users who had an associated data set and some number of taps were included. Users who did not touch the tested button during a performance were excluded — this is because there is no way to be sure that they didn't tap the button because they truly were or were not enjoying the performance, or if they had simply stopped using the system.

This test was run on the group as a whole, and the two groups divided by familiarity. There were no differences in correlation found between the familiar and unfamiliar groups. Statistically significant correlations were found between the number of Enjoyment button taps and the ratings of enjoyment for the experimental performances (whole group results: P1, r_s=.420, p=.002; P2, r_s=.580, p<.0001).

5.5.6 Comparing real-time behaviour to qualitative reports

Finally, I looked into the qualitative data corpus to compare what users did in the moment vs what they reported afterwards. Here there were a number of inconsistencies. Two notable examples were:

1. One respondent who saw Tim's instrument before the concert made tapped the 'enjoyment' button 137 times during Performance 4 (Tim's conventional piece), more than twice as much as any other performance. But, in the qualitative assessment of that performance this participant reported, *It was a bit flat.*

2. A respondent from Group 2, familiar with Dianne's instrument, tapped the 'enjoyment' button 108 times during Performance 4, also more than twice as much as any other performance. In the qualitative feedback they reported, *It seemed a bit disjointed.*

5.5.7 Video data

Links to the video documentation from this study are available in Appendix F.

An an audible click to mark was made during the recording at the point where Metrix was made active. This made it possible to sync the video footage and the real-time data, and analyse of the performance at points of audience agreement about 'enjoyment' and 'error' in the time series data. I was motivated to see whether features of 'enjoyment' and 'error' could be extracted from the video, and Study 1 was the first exploratory use of this method.

The histograms showing the distribution of the button tap data over time are available in Appendix E.

Figure 5.12: Screen shot of video documentation, with real-time 'enjoyment' and 'error' events marked.

Identifying events

Ascertaining where 'enjoyment' and 'error' events was not always straightforward, so I implemented a set of rules that would guide this given the data at hand. These were as follows:

'Enjoyment' events: The threshold for this is a consensus of 6 audience members in a given 5s time bin. The event is considered to have ended when it has gone to a peak, and then is followed by a bin containing a consensus of peak-5, or a consensus that falls below the threshold of 6.

'Error' events: The threshold for error events is a consensus of 2 or more audience members in a given 1s time bin. The event is considered to have ended when it is followed by a bin containing less than the threshold of 2. Contiguous time bins over this threshold are grouped into one event.

The events are marked in the histograms in Appendix E.

Coding the video

In the analysis of the real-time data and the video documentation I developed a method of coding the video, and grouping those codes into themes.

Similar to the rationale described for the analysis of the qualitative data described in Section 5.4, I performed this video analysis on my own in order to make clear the ownership of biases, and because I have the most in-depth knowledge of the subtlety of the data and its questions, I was uniquely positioned to perform this analysis given the video data.

The process I followed was coding, then grouping codes into themes. First, I familiarised myself with the video data by watching it several times. Then I would code the data event by event, by noting what was happening in the video at the times of the indicated events (these were marked in the video). I first coded each 'enjoyment' event, and then each 'error' event, so any connected events or narrative that might connect events of one type would not be lost.

I coded the video footage by recording what I thought, based on my objective observation, was the element, quality or event to which the audience was responding. There were occasional instances where the cause of an event was impossible to determine, and these were indicated in the video coding sheet with a question mark. After coding for one of the event types, I then watched the video again to see if there were any contextual or time-based subtleties that I had missed, and to double-check my impressions of the salient feature of the event.

In this approach to coding the video I considered a number of aspects: The sound, the performative qualities, what the performer was doing, their body language, and events over time.

After coding the video, I grouped these codes into themes. Video coding documentation is in Appendix F.

Features of 'enjoyment' events

The coding of 'enjoyment' events clustered around the following themes:

1. **Theme/compositional:** Times when a theme or musical motif is developing, is established, is resolved, or ends.

2. **Novelty/change:** When a motif or theme changes, or a new motif begins; novelty.

3. **Sound/Musical features:** Passages characterised by rhythm or melody, increasing complexity, or when the performer enhances or adds an element to the existing sound profile.

4. **Performer action/flow:** When the music achieves intensity of rhythm, flow, or when the performer is demonstrably adding to the musical output.

Though all themes were observed in the experimental and conventional performances, there were predominating themes for both styles. In the experimental performances, Category 1 and 2 were the driving forces of enjoyment. In the conventional performances, all categories were observed, but category 4 (Performer action/flow) predominated. This again suggests an audience engagement with the underlying musical language.

Features of 'error' events Audience-indicated 'error' events were, for all performances, less common than 'enjoyment' events. These 'error' events tended to appear as spikes in the histograms with a sharp onset and dropoff, whereas 'enjoyment' events tended to occur far more often but with less localised agreement among the audience.

The 'error' events in the video were also thematically coded, separate from the 'enjoyment' events. The themes that emerged for error were:

1. **Sound and music related:** A wrong pitch or inconsistent rhythm.
2. **Abrupt changes:** Sudden loud sounds, abrupt changes in volume, unexpected elements.
3. **Trivial technical error:** Trivial errors such as the performer hitting the mic stand.
4. **Performer action/reaction:** Facial expressions indicating a mistake, moments of hesitation, moments where they don't seem to be in control, moments where their intention is unclear.

In all performances, errors in themes 2, 3 and 4 were observed. However, for the conventional performances errors were also observed from theme 1 (sound/music related), and these were not present for the experimental performances. This suggests that the audience was able to apply their existing knowledge of musical convention to the former, and this was absent for the latter.

Post-hoc opinions in real time A curious feature of the real-time data was the presence of 'enjoyment' taps at the end of performances, occuring at the end of the music. These were present in performances 1, 3, and 4.

5.6 Discussion

The central question of this study was:

> What is the influence on audience enjoyment of familiarity with the technical aspects of the instrument, and musical style?

5.6.1 The influence of instrument familiarity

The above results suggest that being familiar with the instrument (when this means prior technical knowledge) does not have any appreciable impact on audience ratings of Enjoyment or Interest, or the rank ordering of performances. By extension, I suggest that though transparency as described by Fels et al.[68] may be an important factor for audiences, simply knowing how an instrument works is not sufficient to achieve transparency. This counters the suggestion of a participant in a study by Fyans et al. [75], who suggested that if they knew how the instrument worked they might have liked the performance more. It also suggests that pre-concert demos may be of little use in this regard; if the goal is audience transparency with the goal of greater enjoyment, then demos likely have little effect.

There was one area where familiarity had a significant effect, and this was on the ratings of Understanding between the familiar and unfamiliar groups. The ratings of Understanding were significantly different among the familiar and unfamiliar groups in the post-performance survey, and these ratings were consistent in the post-concert survey. This suggests that the audience's opinion of their understanding was stable over this short amount of time, and also implies that being shown the instrument causes a difference in one's estimation of instrument understanding. However, when asked if they could play the instrument, both groups rated their ability significantly lower, and there was no difference in these ratings between the familiar and unfamiliar groups.

This finding lends nuance to Fels et al.'s notion of transparency [68]. They decscribe transparecny as 'a quality of a mapping ... transparency provides an indication of the psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping.' By this definition, transparency is a deeper and nuanced understanding of an instrument, its physical and psychological aspects, and how it creates sound. For this study, we gave each group a technical tutorial on one instrument, which provides technical familiarity, and the group that was familiar with the instrument rated their Understanding higher. However, when rating their ability to play the instrument — something that would require in-depth physical knowledge, experience, and practice, and would give the audience a deep understanding of the mapping of input to output — both groups rate their ability similarly low. This reinforces the suggestion that technical knowledge of an instrument does not address the confounding factor of a lack of transparency, and that transparency is a much subtler, more nuanced quality.

5.6.2 The influence of musical style

Where familiarity seemed to have no appreciable effect on audience ratings of Enjoyment and Interest, musical style had a demonstrable impact: Conventional performances were ranked significantly higher for the aspects of Enjoyment, Interest and Understanding. Additionally, the conventional performances were significantly preferred in the rank ordering data.

This is an important finding when we consider the differences between the two conventional performances. The performers both performed in a 'conventional' way in the context of their practice, but the performances were still quite different: Dianne's music was folk singer-songwriter, and Tim's was rhythmic techno. The finding of preference for conventional performances is therefore less likely to be simply a preference for a particular type of music, as the performances were still very different, and both significantly higher rated and more preferred.

It is possible, therefore, that the audience is responding not to the musical genre, but to the fact that they had a frame of reference that allowed them to understand the music. Compositional features such as clear structure, melody and rhythm, which were common to both conventional performances, may be the basis of this reference. The qualitative data lends weight to this suggestion. For the conventional performances, compositional and continuous sound features (such as 'flow') were prevalent themes in audience 'likes', and prevalent themes in audience 'dislikes' for the experimental performances.

These findings are not conclusive, and much more study is needed to make conclusive statements about these conditions. This is, however, an interesting avenue for future work.

5.6.3 The larger research questions

As well as the central research question that guided this study, this work also adds to two of the central questions that guide this thesis:

- **RQ1:** With no external stylistic frame of reference, how does the audience perceive error in DMI performance, and how does this affect enjoyment?
- **RQ4:** How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

How audiences perceive error

The qualitative data lends insight into audience perception of error without a frame of reference. As mentioned above, comments about the compositional aspects of the performance, such as the structure and flow, were cited more often as 'likes' for conventional performances, and 'dislikes' for the experimental performances. This may indicate that audiences, while they are receptive to experimental music, do carry with them expectations about what music is, and when these expectations are not met this impacts enjoyment.

This suggestion intersects with Dervin's sense-making methodology, discussed in Chapter 2. Dervin states that, when faced with a gap in knowledge, a user builds a 'bridge' across that gap through a combination of available contextual information, informed primarily through their own contextual knowledge. The qualitative data suggests that this is the case — this audience was clearly using an existing frame of musical reference when watching the experimental performances, and disliked that this available frame of reference was not useful for understanding. Again, this is a preliminary finding, but suggests that Dervin's sense-making methodology may be useful in understanding how audiences perceive DMI performance, the music for which is often extremely experimental and abstract.

The real-time data also has findings that are relevant here. 'Error' events in the experimental performances tended to be related to abrupt changes (suggesting that the audience is trying to perceive a pattern and an error is considered to be an interruption of pattern), as well as trivial technical errors and performer reaction. The latter two themes indicate, again, that it's possible that as well as their existing frame of reference, the contextual information that Dervin mentions also plays a role.

Combining real-time and post-hoc data

In relation to RQ4, there were a number of findings that emerged from the process of finding useful ways to integrate real-time and post-hoc data. Since there is no established methodology that combines post-hoc and real-time data, this study has been an exercise in understanding how to integrate these two data sets.

This study found inconsistencies between the two data sets — there was no profound correlation between the real-time indications of 'enjoyment' and 'error'. This lack of consistency, as demonstrated both qualitatively and quantitatively, suggest that there are differences between what we think during an experience, and how we report it afterwards. This suggests that time may play a role in our opinions, and that as time elapses and we think about an experience, compare it to other experiences, or digest it, our opinions change from what we thought in the moment, acquiring nuance. This is not to say that real-time data is 'true' and survey data is not; more likely, this suggests that our musical opinions are not fixed and finite.

There is also an question of why was there was not an increase in 'enjoyment' tap rates for the conventional performances, in line with the increased ratings of Enjoyment. One possibility is that people may be less likely to tap 'enjoyment' buttons while they are engrossed in an enjoyable experience, or perhaps this is because our opinions are formed over time, and enjoyment in a reflective sense does not necessarily mean that we indicate more enjoyable moments in real time.

The curious appearance of the end-of-performance 'enjoyment' events also bears consideration. As these events occurred as the music was ending or after it was over, these could be considered as a the registering of opinion on what the viewer has just seen. These end-enjoyment taps are perhaps more akin to a reflective 'I enjoyed that' as opposed to the in-the-moment

‘I am enjoying this’, or perhaps a registration of appreciation. It is also equally possible that while applauding the audience unknowingly registers button taps. Without further study no specific inferences are possible, but it certainly presents an interesting phenomenon for further investigation.

5.6.4 Other findings

Different novelty effects

The novelty of a new instrument did not seem to be a salient feature of the qualitative data past the first performance. Further, since the first performance was middle-ranked, it appears not to have profoundly impacted the ratings, either positively or negatively. This suggests that although audiences do notice newness of an instrument, it is not a distracting or overwhelming factor.

Though instrument novelty did not seem to feature prominently, novelty in music did. Novelty was found to be a theme of ‘enjoyment’ events, particularly for the experimental performances; for these performances, the beginning of a new motif or a new additional sound or effect was a prevalent feature of ‘enjoyment’ events. As discussed above, this may be because the audience is seeking pattern, but any suggestions on this cannot be made without further study.

Clues about the nature of error and enjoyment from the video data

Perhaps the most important finding is that the real-time data suggests that ‘enjoyment’ and ‘error’ are not binary terms, and are experienced differently. The ‘error’ events did not occur opposite the ‘enjoyment’ events, and ‘error’ indications were much less common overall. Further, ‘enjoyment’ seems to have a cumulative effect that grows to a peak over time, whereas ‘error’ events tend to be spikes of agreement that drop off straight away. This is made more intriguing by the fact that the most enjoyed performances did not have fewer ‘error’ indications. This suggests that ‘error’ in some way contributes to enjoyment, or is an aspect of enjoyment. More study is needed, however, to make conclusive findings.

Another notable feature of the real-time data was that the highest-rated performances were not the most error-free.

The analysis of the video documentation revealed themes of error and enjoyment events. Contrary to the suggestion by Fyans et al. [75], this audience *did* identify errors in experimental DMI performance, but the nature of those errors is different from DMI performances in a conventional style.

Audiences in this study seem to rely on contextual information to deduce error, such as performer action or inaction. For the conventional performances these were also factors, but musical inconsistencies were much more prevalent. This is evidence that audiences are forming their judgements of error based on musical vernacular knowledge, but when that is absent, they are relying on contextual knowledge to make sense of what they're seeing. This is akin to Dervin's sense-making methodology [55] described in Chapter 2, which posits that, when audiences are met with a gap in their knowledge, they rely on their own frame of reference and contextual clues to make sense of it. With the conventional performances, there was still reliance on existing frames of reference (in this case musical vernacular) to identify error.

5.6.5 Limitations

This study, and by extension its results, has limitations.

This is a study of the perception of four performances, by two performers, on two instruments, evaluated by one audience. As such, these results should not be widely generalised to all audiences, or even to audiences of DMI performance. The post-structuralist viewpoint mentioned in Chapter 3, therefore, is useful here, as there is no guarantee that these or any results would be reliably consistent for all audiences and all contexts.

That said, the results of this study do illuminate some interesting insights and directions for future work. Though this methodology and these findings are in their infancy and as such cannot and should not be generalised to all audiences, or even to general audiences of DMI performance, they do indicate areas that will, potentially, add to a better understanding of audiences, whether those future results confirm or refute the findiings in this chapter. The results that arose during this study from the combination of the real-time feedback and the video documentation are of particular interest and may prove useful, but this is only the first test of this technique; for example, though thematic analysis of video codes has presented insights, this may not be the best or only way of querying this data.

Finally, this study is not intended to demonstrate that one musical style is 'better' than another, or even to suggest that self-reported audience enjoyment ought to be the guiding factor in the DMI design process. Different musical works engage the audience in different ways, and some of the most profound musical experiences are also the most challenging in the moment. We do, however, suggest that these findings provide valuable insight into the audience experience of novel DMI performance, and that time spent making the audience understand the technology is misplaced. Though many DMIs are often tightly connected to the music they are used to produce, there is value in considering which aspects of audience experience are influenced by technology, and which by aesthetic and stylistic factors.

5.7 Summary

This chapter presented the first study of this thesis, which investigated the relative effects of instrument familiarity and musical style on audience perception of DMI performance.

In the context of an evening concert, this was investigated through a study that involved two musicians, each of whom play a novel, self-built DMI. The audience was split into two, and each half was given a technical tutorial on one of the instruments, creating a contrast in familiarity. The performers then played two pieces, one in an experimental style, and one in a conventional style. Post-hoc audience data was collected through written surveys as well as real-time audience indications of 'enjoyment' and 'error' via a mobile phone interface.

The central research question for this study was: *What is the relative influence of familiarity with the instrument and musical style on audience perception of DMI performance?* Contrary to the suggestion of participants in Fyans et al.'s study [75], we found that familiarity with the instrument had no impact on audience interest or enjoyment of these performances novel DMIs. However, the difference between conventional and experimental musical styles had a significant impact on audience perception.

Further, this study offers insight into the limits of transparency in DMI design [68], showing that insofar as transparency is important to the audience experience, it cannot be addressed simply by explaining the instrument or imparting technical knowledge. Though this study provides insight into the role of transparency, it does not directly address the question of whether an intuitively obvious relationship between gesture and sound improves the audience experience. Gurevich and Fyans's work comparing the theremin and Tilt-Synth [87] provides some hints in that direction, but this work on transparency and the audience is in its first stages, and further studies are needed to confirm how transparency is achieved and whether it has a meaningful effect on audience enjoyment, as well as to understand the design choices that support it.

This study also lends insight into two of the larger research questions guiding this thesis, RQ1[5] and RQ4[6]. For RQ1, the real-time data showed that error and enjoyment were not perceived as opposite terms by the audience: The audience did not only indicate errors when they were not indicating enjoyment, and the tap behaviour on the error and enjoyment buttons were very different, with the enjoyment button being used far more often and enjoyment appearing as a cumulative effect over time, whereas error tended to have a sudden onset and sharp drop-off. Further, being familiar with the instrument did not have a significant effect on the audience's perception of enjoyment or error.

In combining the post-hoc and real-time data, it was found that the most-enjoyed performance also had similar rates of error indications as the lower-rated performances, suggesting that error is not viewed as uniquely negative, or at least that errors do not render a performance unenjoyable. Further, the post-hoc data of prolific users of the buttons was queried, and their post-hoc comments did not match their real-time behaviours. In this way, there may be that the passage of time, or perhaps the synthesising of opinion into language, has an influence on our perception of DMI performance.

I also stress that although these results are demonstrable and may lend insight into larger questions and illuminate future areas of study, these effects they may not extend uniformly to all performers and all instruments, or all audiences in all contexts.

[5]RQ1: WWith no external stylistic frame of reference, how does the audience perceive error in DMI performance, and how does this affect enjoyment?

[6]RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

5.7.1 Carry forward

Through its central question and the combined approach to data gathering, this study added valuable insight to three of the research questions guiding this thesis. Additionally, I found useful strategies to examining and understanding the real-time data, as well as combining it with the post-hoc results. With this methodology solidified, I was able to apply it to the subsequent studies, found in Chapter 7 and 8.

For these studies, I created the experimental instruments. Before describing these studies and their results, I use the following chapter, Chapter 6 to describe the design process undertaken to create these instruments.

Chapter 6

The experimental instruments

This chapter describes the two instruments that were designed and produced for this research: MOAI and Keppi.

The process of instrument design is complex, labour-intensive, and requires expertise in multiple domains, including sound design, interaction design, and fabrication. Most importantly, though, this work demands an engagement with what is is to make an object that Gell terms a ‘technology of enchantment’ [79] — that is, an object that is used for artistic means.

This chapter first outlines the rationale for undertaking this complex task in Section 6.1. Next, in Section 6.2 I outline the set of design values with which I approached the design of these instruments. Finally, in Sections 6.3 and 6.4 I describe how these design values were applied to each of the experimental DMIs, and how each of these instruments met the needs of the study for which they were produced.

6.1 Rationale

Engaging in the process of DMI design for the experimental offered distinct advantages to this research that justified this investment of time and effort.

Firstly, designing the instruments allowed me to ask precise questions in this study. I learned, while searching for musicians to play Study 1, that it is often extremely challenging, if not sometimes impossible, to find the right performer with the right instrument to fulfil the needs of a particular line of inquiry. By designing the instruments, I was able to design instruments that supported the goals of my research, instead of having to comprimise the central questions because I could not find suitable instruments.

Secondly, designing an instrument in more than one version allowed for a direct axis of comparison of audience perception. DMIs are notoriously diverse in their physical design, functionality and interaction. As a result, When comparing audience perception of two DMIs it becomes difficult to attribute any difference in audience response to the differences between two given DMIs, and to rule out the many competing and complex factors at play (such control dislocation, transparency, and so on). Although designing versions of instruments does not eliminate this complexity entirely, it does provide a credible way to compare audience response.

Finally, producing the instruments and comparing those versions allows this thesis to offer insight into useful design strategies. I have been involved in interaction design as a practice for most of my career, particularly the design of usable art objects and audience-focused experiences and installations. This means that I have brought to this process a deep knowledge of audiences, along with extensive knowledge of materials, object design, and production. In this way, I offer insights gained from an approach to DMI design that prioritises the audience experience, as well as the viability of the DMI as a creative tool.

6.2 Values, approaches, and considerations

DMI designers are confronted with a huge amount of freedom. Because the sound of a DMI is not necessarily connected to its materials, physical form, or the gestures used to play it (see discussion of 'control dislocation' in Chapter 2), there are very few constraints on what a DMI looks like or the materials it's made from.

This freedom brings with it a huge creative opportunity, but at the same time a lack of constraints means that there are few external influences to guide crucial design decisions, and certainly not the strict limitations faced by traditional instrument designers who must resolve design with the physical capabilities of the instrument. There is also a lack of cohesive design values within the DMI community, stemming directly from the value placed on experimentation and exploration.

To navigate this enormous freedom, I constructed a set of design values that guided the production of these two experimental DMIs.

6.2.1 Privileging of craft and 'mid-diversity'

DMI design has in the past been called 'a highly specialized field of HCI' [204]. Though it is certainly a branch of interface design, contextualising it with respect to HCI does not capture the complex, competing and contradictory considerations that are outside those inherent in most HCI design projects. It is therefore useful to adopt a better term for the craft and design of instruments, and Jordà in 2004 offered the term *digital lutherie*:

> *Digital lutherie is in many respects very similar to music creation. It involves a great deal of different know-how and many technical and technological issues. However, like in music, there are no inviolable laws. That is to say that digital lutherie should not be considered as a science, but as a sort of craftsmanship that sometimes may produce a work of art, no less than music.* [107]

This statement expands upon Cook's 2001 assertion, that 'Musical interface construction proceeds as more art than science, and possibly this is the only way that it can be done' [47]. Though the suggestion by both Jordà and Cook, that there exists no prescriptive way of going about this process, is undoubtedly true, the artistic, exploratory, material-based nature of instrument design does not preclude this process from being articulated, or from being purposefully undertaken.

There is currently a deficit of language to describe the *process* of DMI design within the NIME and HCI communities. Though much as been written about the process of DMI design, this literature is largely focused on the technical aspects, and not the experience-based process of instrument crafting. Despite this a lack of process language, there exists a plethora of terms and descriptors for the *functional* aspects of instruments. Because of this focus on functional aspects and a de-emphasis of the process of instrument design, this exploratory process still remains largely undocumented, dismissed as 'artistic' and not scientifically relevant. (There are concerted efforts to change this practice, most recently by Armitage et al. within the NIME community [6] which builds on the established work by Sarah Kettley in the greater digital craft domain [116], but this shift is, at the time of writing, still in its infancy.)

The fundamental goals of HCI are not necessarily applicable to DMI design. For example, though usability is a primary goal of HCI, Donald Norman indicates that musical instruments do not necessarily benefit from this as an end goal:

> *Usability is a complex topic. A product that does what is required, and is understandable, may still not be usable. Thus, guitars and violins do their assigned tasks well (that is, create music), they are quite simple to understand, but they are very difficult to use. The same is true of a piano, a deceptively simple-looking instrument. Musical instruments take years of dedicated practice to be used properly, and even then, error and poor performance are common among nonprofessionals. The relative unusability of musical instruments is accepted, in part because we know of no other alternative, in part because the results are so worthwhile.* [151, p. 77-8]

For this reason, I instead used Jordà's concept of micro-, mid- and macro-diversity as a starting point. This taxonomy considers the levels of diversity of playing that are possible when interacting with a DMI: The tiny nuances and subtleties of playing (micro), the contrasts in performance (mid), and the ability of an instrument to be played in different musical styles (macro) [107].

All of these levels of diversity are important to consider, but for these experimental instruments I was interested particularly in the aspects of mid-diversity, as the studies aimed to compare audience perception of performances.

6.2.2 Functionalism: Form follows MAYA

In order to guide the physical design of these instruments, I drew on two concepts: Functionalism, and Raymond Loewy's concept of MAYA (Most Advanced, Yet Acceptable).

Though it emerged from architectural theory, Functionalism has been deeply influential on a wide range of design thinking in the 20th century.

The term's origins stretch back far further than the 20th century, but it is an 1896 magazine article in which architect Louis H. Sullivan is considered to have coined the following phrase:

> *It is the pervading law of all things organic and inorganic, of all things physical and metaphysical, of all things human and all things superhuman, of all true manifestations of the head, of the heart, of the soul, that the life is recognizable in its expression, that* ***form ever follows function. This is the law.*** [186]

A rigorous exploration of the usefulness of Functionalism as a point of departure in DMI design is outside the scope of this thesis, but it is useful to briefly describe the term's origins and applications here. Though it was originally applied to buildings, Functionalism's approach of constructing form based on the end object's function was a direct result of advances in building materials emerging in the 20th century. With materials such as reinforced steel and concrete, buildings were no longer limited in height or size by the materials used to make them. Now that so much was possible and buildings could take any form, what should that form be?

This situation is generally analgous to the freedom that control dislocation has afforded DMI designers. A strategy that architects employed to confront this — particularly those within the Bauhaus from 1919 to 1933, but this inquiry was occuring on a larger scale — was to establish rules and approaches to underpin the use of new materials. Functionalism was born specifically out of this discourse, and in its most basic sense, dictates that if a building can look like anything, then **what it is used for should dictate what it looks like**.

This is useful for DMI design. Control dislocation is a major confounding factor for audiences, and Functionalism is useful in this context as it supports the intuitive link between action and sound that is often missing. However, a Functionalist approach does not necessarily create an instrument that is also interesting for performers to interact with and play, and becomes cloudy when we consider the 'function' of a computer inside the instrument, which is extremely multifaceted. Because I could not see a way that a strict Functionalist approach would support the process of designing in interesting tools for artistic creation, and because the presence of a computer has to be carefully navigated, I augmented this Functionalist approach with a related concept from Raymond Loewy, the influential industrial designer from the mid-20th century.

Loewy was a designer of mass-market consumer products through the 1950s, 60s and onward, an age of rapid technological development. Loewy adopted a strategy he termed 'MAYA', or 'Most Advanced, Yet Acceptable' [128, quoted]. Loewy's anecdotal advice was that consumers will be most receptive to objects that are familiar, but have novel aspects.

Thinking along Formalist and MAYA lines led me to limit these designs to one set of DMI performers. I chose percussionists, both because I have a personal interest in percussion and percussive instruments, and because percussion closely couples the action-sound relationship. This was a straightforward way to make an intuitive connection for audiences, and MAYA meant that, if I kept these instruments within the realm of percussion, I could also introduce interaction elements that were not necessarily in the realm of percussion but that percussionists would respond to and audiences would intutively understand (such as Keppi's three dimensionality, which is discussed in Section 6.4).

6.2.3 Interactive pluralism

Constraint in DMI design can serve two purposes: To support the intuitive relationship between action and sound by eliminating competing factors, and to encourage the musician's personal playing style.

For the audience, I used constraint in the design of the DMI to declutter the relationship between action and sound. In the case of MOAI, this was by making the only sound-producing interaction one of striking; there is a direct and established link between input and output. Further, constraint was applied to the sound design. Both instruments had limited sound palettes, specifically so there was no complex synthesis or complicated mapping that would further complicate the action to sound relationship.

Constraint's most useful application, however, was when considering the instrument's interactive possibilities. Though the performer's experience is outside the scope of this thesis and was not evaluated for that reason, it was still important to this research that these be musical instruments that were artistically useful (see the discussion of complexity in Chapter 3), and constraint with the goal of designing affordances that would be artistically useful.

Gurevich's assertion that constraint is a way of encouraging personal style is compelling [88], and its further confirmation by Zappi and McPherson [214] with respect to appropriating behaviours makes it more so. Jordà's identification of a range of diversity within interctions [109], categorised as micro-, mid- and macro-diversity, intersects with this notion of style, as he states that it is in the mid-diversity range that refers to the distinction between multiple performances on the same instrument.

Though the recommendation of constraint and this diversity taxonomy are extremely useful for separating the interaction issues at play, neither lend any practical insight into what a usefully constrained affordance might look like. For this I looked further than the DMI domain to other research fields. Turkle and Papert's concept of epistimological pluralism [195] was particularly useful here, despite being from a tangential field (the original context of this work was gender and computers).

Episitomological pluralism is defined by Turkle and Papert as 'the validity of multiple ways of knowing and thinking' [195], or the idea that there is not only one way to learn, think about, or use things (in their case, technology). Further, they state that knowledge acquired and applied in disparate fields is often innovative, because it breaks with the established ways of doing things.

When this idea is applied to affordances of a DMI, personal style could also be described as a *plurality of interactive methods* in which the musician might engage with an instrument. As such, I wanted to design affordances that were constrained, but offered a range of interactive pluralities. In the case of Keppi, the constraint was that it could only be played by tapping one of four electrodes on the instrument. However, the plurality of interaction was applied in the fact that the instrument was round, hand-sized, and needed to be moved in three dimensions. This meant that there was a large range of ways that this interaction could be performed, guided by the player's stylistic choices.

This pluralism intersects with Gaver et al.'s concept of ambiguity as a resource for design, particularly their identification of ambiguity as it relates to relationship [77]. They state that this kind of ambiguity 'creates the condition for a deeply personal projection of imagination and values onto a design', which is a guiding principle to designing for a plurality of interactions.

6.2.4 Summary

In order to navigate the boundless freedom inherent in DMI design, I adopted a three-pronged approach.

Firstly, I considered this a craft-based task, and therefore focused on the craft-based aspects instead of established HCI approaches to design. I concentrated on creating what Jordà terms *mid-diveristy* in order to create contrast between the performances. Focusing on mid-diversity supports the goals of this research, as I was examining the differences in audience perception of different performances on one instrument.

Secondly, I looked to the concept of Formalism, combined with Loewy's idea of 'Most Advanced, Yet Acceptable', to guide the design of the physical aspects of these instruments. This was to further mitigate control dislocation for the audience, yet still present interfaces to performers that were artistically interesting and useful.

Finally, I considered the goal of 'interactive pluralism', based on Turkle and Papert's ideas of epistomological pluralism [195]. This value, when applied to the constrained affordances of a DMI (as recommended by Gurevich et al. [88]), places emphasis on developing a diversity of methods of interact with that affordance. This is related to ideas around ambiguity in design as it relates to ambiguity of relationship, proposed by Gaver et al. [77]. I reflect on the effectiveness of this approach at the end of this chapter, in Section 6.5.

6.3 MOAI

MOAI was an instrument designed for Study 2, which investigated the impact of gesture size on audience perception of DMI performance, and queries whether the DMI's physicality can influence the gestures used to play it.

The motivation of this study and existing work related to it are fully explored in Chapter 7, but I will summarise the central question here. Gesture is a primary area of interest in DMI research, and the community has produced a huge amount of literature on the topic. In HCI gesture is also a major area of inquiry, and a range of models for understanding how audiences perceive gesture exist.

But, in these two disciplines there is no work that creates a comparative case between two physical repertoires. This work was motivated by the desire to understand this more precisely, by comparing audience perception of the same DMI at two contrasting physical sizes.

To provide a comparative case, this interface was produced in two sizes: One small, and the other 3.5 times larger. The internal hardware, the physical proportions, and the functionality were identical.

This section details the design and production of MOAI, by detailing first the goals of the study and the influence of the musicians it was designed for. I then describe the sensor design and programming, the sound design, the interaction design, and the fabrication of the instrument, and how the design values described in Section 6.2 influenced these.

6.3.1 Study 2: Examining gesture

Gesture is a heavily discussed topic in DMI research [105], and general consensus that gesture is important. Despite this agreement, there is little indication of what a gesture that is effective for the audience might look like, how the DMI design might support this.

I wanted to be able to compare the impact of different scales of the same gesture on audience enjoyment. For this I needed an instrument that privileged two performance aspects: Gesture interaction, and visibility of that interaction.

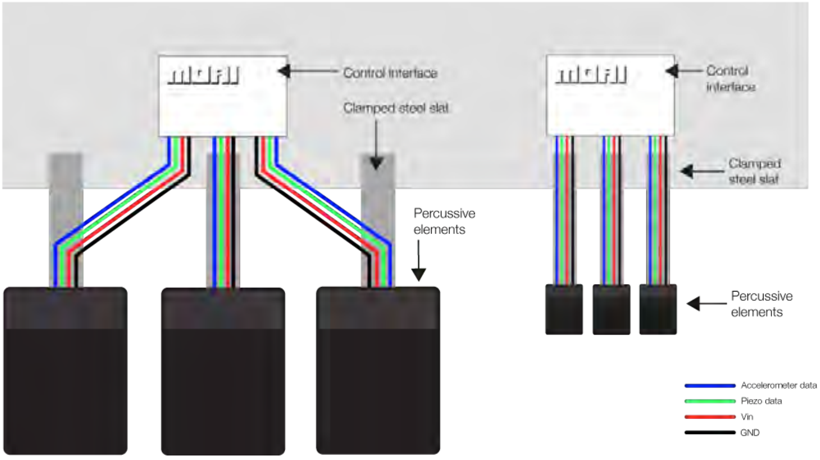
As the cause and effect nature of percussion addresses some of the confounding complexity (as discussed in Chapter 3), this would be percussion instrument. In order to change the size of the gestures used to play it, I decided to build one instrument in two versions that were identical in hardware, function, sound and interaction, but only differed in terms of scale. In this way, I could directly compare the data and see if design factors had an impact on audience opinion, thereby gaining insight directly into this problem instead of having to isolate the relative impacts of the different aspects of two different instruments.

MOAI was produced specifically for this purpose. It is played with percussive strikes, eliciting gesture, and the scale of the instrument would make the differences in that gesture visible.

6.3.2 MOAI: Technical details

A technical diagram for MOAI is available in Appendix G.

Figure 6.1: Diagrams of the setup of both MOAI versions. Left: Large. Right: Small.



MOAI is an acronym for Multiply Oscillating Actuated Interface. MOAI consists of three percussion elements that each sit on a strip of steel that is clamped to a table, so they bounce in response to being hit. All 3 elements are connected to a central interface that contains the computer (which continually processes the three streams of sensor data, one from each element). The control interface has a few basic system controls (mute, shutdown, volume, and potentiometer for controlling the overall sensitivity of the instrument). See Figure 6.1. The percussive elements are identical in materiality, proportion, and internal hardware, and differ only in terms of scale — one version is large, with each element measuring approximately 30cm x 40cm x 20cm, and one is 3.5 times smaller, with each element measuring 12cm x 15cm x 5cm (see Figure 6.2).

The percussive elements each sit on a steel slat that is clamped to a table. These steel slats are flexible, and cause the elements to bounce up and down as they are struck. This further reinforces visibility, as the elements oscillate in response to being played.

The hardware inside each MOAI element is identical:

- Piezo network to sense the velocity of strikes
- An accelerometer, to sense the movement of the box

The function of these components in MOAI is discussed further below.

Inside the control interface is a Bela[1] board that processes all six streams of analogue data, and produces the sound output in response.

6.3.3 The musicians: Ex Easter Island Head

This instrument was designed with input from the participating musician, Benjamin Duvall of Liverpool-based percussion ensemble Ex Easter Island Head[2]. The instrument's name, MOAI, is an acronym standing for Multiply Oscillating, Articulated Interface. This also references the moai, which are the giant head sculptures on Easter Island (see Figure 6.3).

The band has gained notoriety in the UK music scene for their percussion performances, in which they use open tuned electric guitars as percussion surfaces (Figure 6.4), which they strike with a variety of mallets and sticks. The tonal qualities of the ringing guitar strings give their performances a droning and meditative nature.

[1] https://bela.io
[2] `http://exeasterislandhead.com`

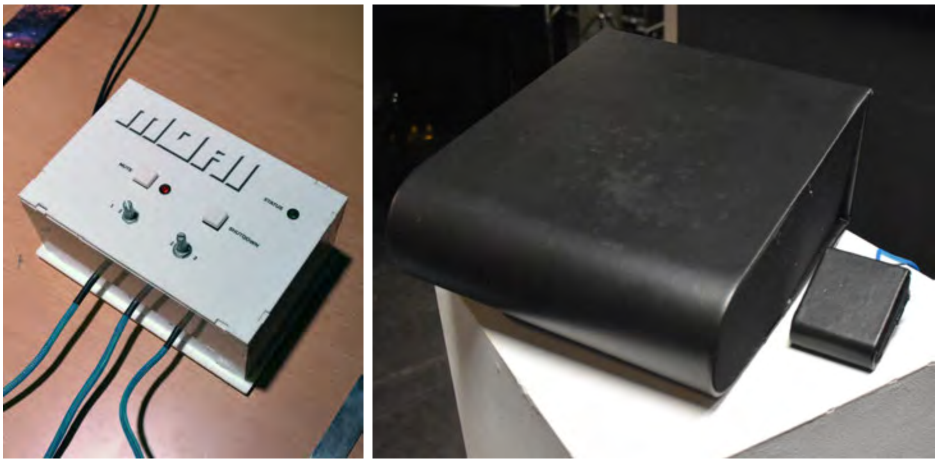


Figure 6.2: Left: The MOAI control interface. Right: A large MOAI element and a small MOAI element, illustrating the difference in size (Large, 40x30x20cm; small, 12x10x5cm).

Figure 6.3: A moai on Easter Island. [7]

6.3.4 Design process

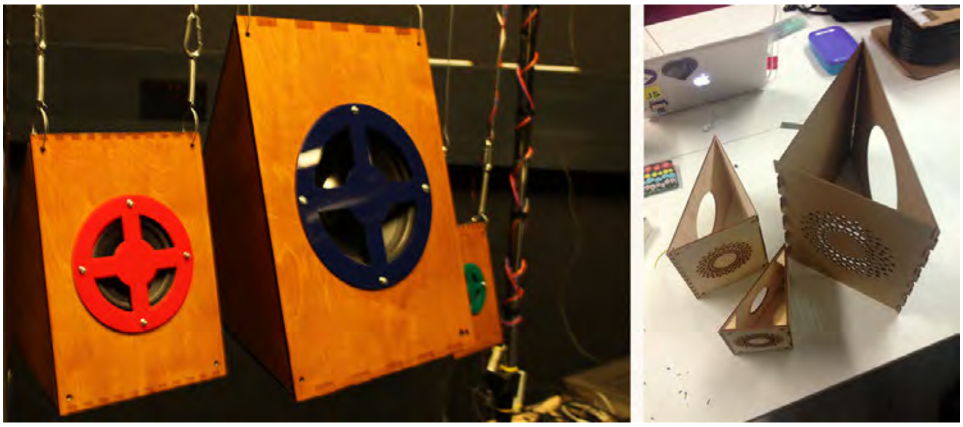
Point of departure, interaction, materials

MOAI was inspired by B.U.R.T., an instrument I had built in my first year in collaboration with Robin Rimbaud aka Scanner[3]. It consisted of wooden boxes that hung on springs (see Figure 6.5). This collaboration was the subject of the Advanced Placement Project that took place during the first

[3] http://scannerdot.com

Figure 6.4: Still from Ex Easter Island Head's video, Six Sticks. [63]

Figure 6.5: Left: Finished B.U.R.T. boxes. Right: Side view of the B.U.R.T. boxes



year of my PhD study, and was the first time I was able to explore how my existing interaction design skills could apply to DMI design.

I had spent a considerable amount of time on the construction and proportions of the boxes, and wondered if there was a way I could use this shape again. I attached one to a piece of steel I found in the workshop I was working in, tested it, and discovered the bounce in response to hit was a very interesting dynamic element (see Figure 6.6). This general shape and behaviour was the point of departure for MOAI.

As I played it I noticed that bouncing the elements was interesting, but so was the ability to stop the bouncing. For this reason I adjusted the box design to have a rounded front in order to lend itself to grabbing with the hand, and incorporated this bouncing/grabbing into the musical interaction.

I specifically designed each MOAI element so it had only one playable surface, and that it oscillated in only one direction and at one frequency (determined by where the steel strip was clamped to the table and the length of the overhang over the table's edge). In this way, all plurality of interaction was focused into how the player started and stopped the oscillation, and the manner in which they made their strikes.

Physical form factor and materials

Each version of MOAI consists of three identical percussion elements, and the contrast in size between the two versions was of primary importance as this would provide the point of comparison in the study. How large was ‘large’ and how small was ‘small’ was dependent on the hardware and fabrication process. For the small version, I made it as small as I could while still containing the requisite hardware. For the large version, I was constrained by the size of the laser cutter bed to cut the largest parts of the housing.

The material choices for MOAI were made by considering their function. They had to be sturdy, so I used plywood for the internal support. The grabbing surface was made by cutting a living hinge into the plywood, but was fragile. To support it, I created internal rounded supports out of cardboard.

Two functional aspects were mass, and that the boxes had to be able to be opened to be attached to the screws that held them to steel slats with a nut and bolt. I had to consider weight at every turn to achieve a bounce on the steel slats. Because of this, and to make sides that could be taken off and re-attached, I made the sides out of cardboard that had been covered in lightweight black fabric.

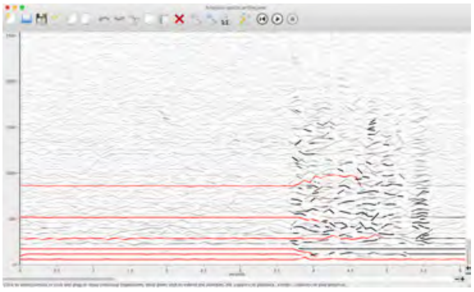
I chose a black vinyl to cover the elements. Materials have a language of their own, and I wanted MOAI to communicate primarily through form and interaction instead of looking like a plywood ‘maker’ project (thereby not communicating as a musical instrument). I also wanted to avoid any communication of wood to avoid a connotation of acoustic instruments.

Sound

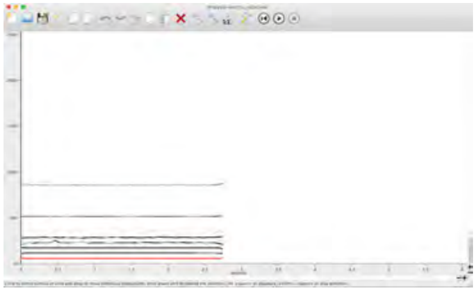
The core sound of the MOAI system was created from the NASA recording of the moon landing, where Neil Armstrong climbs off the space shuttle and

Figure 6.6: Still from video of first test of B.U.R.T. box that would be redesigned as MOAI

Figure 6.7: Screen captures of the audio synthesis process in SPEAR.



(a) Audio file in SPEAR. Relevant partials are shown highlighted.



(b) Extracted partials isolated for resynthesis. 60Hz partial highlighted.

says ‘That’s one small step for man, one giant leap for mankind.’ [149]

The first few seconds of this recording have a mechanical hum in the background. I was interested in isolating this hum as it was rich and brought to mind the drone qualities of Ex-Easter Island Head’s use of tuned guitars.

For this sound synthesis I used SPEAR (Sinusoidal Partial Editing Analysis and Resynthesis) [119]. SPEAR is free software that resynthesizes audio input by representing that input through many individual sinusoidal tracks, or partials. Each partial corresponds to a single sinusoidal wave with time varying frequency and amplitude [120]. I isolated the 60Hz hum and the complementary harmonics I was interested in and synthesised a sample (See Figure 6.7). I then gave this sample a percussive quality with the addition of an envelope.

This sound was applied to the centre box, Box 2. This sound was transposed down a minor third and applied to Box 1 (left), and up a minor third and applied to Box 3.

Each MOAI box produces three sounds:

- The base tone, described above
- The base tone with upper partials
- A sound with low partials

The high partials added to each element’s base tone were created using a recording of percussion on the upper strings of a grand piano. These partials were then added to each box’s base tone (though the partials themselves were not transposed).

For the low partials, I was inspired by the richness of the bass tones of The Sinking Belle, a collaboaration by Sunn o))), Boris, and Jesse Sykes and the Sweet Hereafter [187]. Using SPEAR I analysed this track, and I found a combination of sinusoidal partials that I liked. I then synthesised a sample by combining sinusoidal partials at the relevant frequencies (directly synthesizing from the track wasn't possible as the partials didn't have the amplitude, clean tone or duration I was interested in). This low tone was applied to Box 2, then transposed for boxes 1 and 3 as before.

6.3.5 Sensor design and programming

Inside MOAI's control interface is a Bela [137][4] unit that continually processes the analogue data from the three elements. The software and samples for MOAI are publicly available on Github[5]. The hardware for the large and small versions were identical, except that a network of multiple small piezo discs were used for the small version.

MOAI plays one of three samples — base tone, tone with high partials, or tone with low partials — in response to strikes. MOAI begins playing a sample when it detects a strike on one of the elements, but it determines which sample to play based on the movement state of that box, determined through the accelerometer data.

To detect piezo strikes, the control interface constantly samples the signal generated by each element's embedded piezo network (see Fig 6.8). The piezo input is used to detect strike events, and assign a velocity to the strike. The piezo input undergoes considerable conditioning each time it is sampled (Bela's analogue sample rate is 22.05kHz). First, it is full wave rectified and a DC blocking filter applied.

When any piezo velocity is sensed above a threshold (this threshold is applied to reject noise), Bela looks both forward and backward by 5ms of input to find the highest signal peak. This is because I found that there was no guarantee that the strike that is the first detected would be the highest, so this process was implemented to make sure the highest velocity is applied. This did method did not add any noticeable latency.

The sound that is played in response to a strike is determined by accelerometer data. The accelerometer data from each element is also pro-

[4]https://bela.io
[5]https://github.com/disastrid/MOAI_2016

Figure 6.8: A piezo network in a large MOAI. Piezo discs are 50mm in diameter, connected in parallel, and the solider joints are insulated and protected with hot glue.

cessed continually, the movement state determined by this data indicates which sample will be played. The velocity input detected by the piezo network determines the velocity assigned to that sample.

The sound-producing states determined by box movement are are *moving*, *still*, or *stopped suddenly*. Moving is by far the most common state as the boxes bounce when hit, and continue to bounce after. An element that is moving plays the element's base tone in response to a strike. Starting from a still position, as determined by accelerometer data, produces the element's base tone with added low partials. Striking the box with sufficient velocity produces the element's base tone with added high partials. The threshold of velocity needed for this tone is adjustable using a potentiometer on the control box, to allow for variations in strike velocity between performers.

6.4 Keppi

Keppi was the instrument I designed and produced for Study 3. In this section I describe the motivation for the study that used Keppi and how Keppi's design supported these experimental goals. I also describe the design process, material considerations, electronics and sensor design and processing, and finally explain the design of the behaviour that formed the test condition for Study 3.

6.4.1 Study 3: Exploring disfluency

Study 3 explored the effect of disfluency on audience perception of DMI performance. Disfluency is defined as 'the experience of processing difficulty' [192], and has been shown to result in heightened cognitive processing.

I wanted to design an instrument that would lend itself to being used in a risky way, and was interested in exploring how a percussion instrument could support this. Because of the physical nature of percussion I knew that the players would have excellent coordination and motor skills, so I decided to base the risk factor around them interacting with an instrument that was not a surface to hit, but an object that had to be struck.

For this study I designed an instrument called Keppi[6], a percussion instrument in the form of a cylinder about 65cm long and 12cm in diameter. It has four large electrodes that wrap around the body of the instrument, interspersed with five rows of LED lights, and speakers embedded in either end of the cylinder (see Figure 6.9). When the performer taps one of the four electrodes, the embedded computer receives a trigger via capacitive touch and triggers a sample, applying a velocity based on the signal generated by the piezo network underneath the electrode that was tapped.

I was interested in introducing disfluency into Keppi not as a design feature, but as a behaviour, and was interested in this disfluency being visible to the audience. Six Keppis were produced for this study and all were visually identical, but differed in one behavioural aspect: Some included an accelerometer, and the embedded system constantly monitored the quantity of movement of the instrument. If it was not sufficiently moved through space, the five rows of lights would begin to tick down, ticking up again if Keppi sensed enough movement through space. If not moved enough the lights would tick down, and when they went off Keppi would cease to make sound. Moving Keppi charged the lights up again.

I then produced six identical versions of Keppi, one for each of the performers. and each had one of three behaviour conditions, and a total of two of each behaviour type:

Control group: No disfluency. There was no requirement to move the instrument, and all lights stayed on at all times.

Condition 1: The instrument's lights would tick down at the rate of 1500ms.

[6]From the Finnish for 'stick'.

Condition 2: The instrument's lights would tick down at the rate of 650ms.

6.4.2 The percussionists

For this study I focused on recruiting percussionists with five or more years of training and performance experience, to ensure that all percussionists would be sufficiently skilled and no performance could be dismissed out of hand simply because the performer wasn't very good.

I was fortunate to be introduced to a network of percussionists who were alumni of the Guildhall music program. Four of the participants were from this network. The other two were percussionists with the appropriate level of skill and experience recruited from elsewhere. None of these percussionists usually played DMIs such as MOAI.

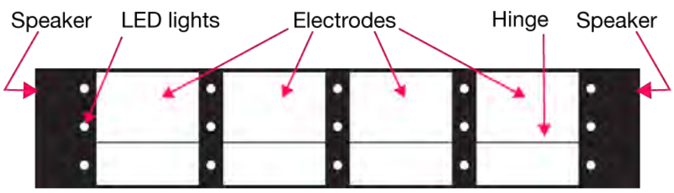
6.4.3 Physical design

Keppi is a tube that is 62cm long and 12cm in diameter. The electrodes begin 12cm from each end and are approximately 7cm wide, with the borders between them approximately 2.5cm wide. See Figure 6.10.

Since Keppi is entirely wireless to encourage movement in three dimensions, it is split in two halves and held closed with velcro strips that wrap around the ends, and the second and fourth row of LEDs. Since this velcro is narrow, black and has a low profile, it blends in with the rows of LEDs. Using velcro closes Keppi securely, but also allows it to be easily opened and closed for turning it on and off or to adjust the components inside if needed.

Figure 6.9: Keppi, with the lights on.

Figure 6.10: Diagram of Keppi and the placement of lights and electrodes.



Considering the player's body and incorporating curve were important aspects of MOAI, and these aspects influenced the physical design of Keppi. Keppi's particular physical dimensions are a result of my desire to make something big enough that would encourage the player to consider their body in relation to it, and not appear to be a tool or toy. I also wanted the instrument to be visible to the audience from the stage. I chose a tube because I wanted Keppi to lend itself to be gripped with a hand, but I chose a wider tube to make it less of a baton and more of a cylinder.

LED lights

The LED lights were an essential and carefully-considered design element. Because of the audience-focused nature of this work, I wanted to build in a visual aspect that meant the audience would be aware of the time element of the disfluent behaviour.

The LEDs used were 3mm, and there were 6 in each of the 5 rows. The lights in each row were equally spaced around the outside of the instrument. This was to ensure that the lights would be visible at every angle, but would not act as a single band of lights (which I was concerned would resemble a thermometer or some other kind of sensor technology).

Consideration of affordances

The ways that a DMI's form factor can act as an affordance is a particular interest of mine, and I was therefore interested in creating a DMI that was not immediately conducive to being placed on a table. I wanted Keppi to be something that encouraged physical interaction, and therefore made it something that was meant to be held.

The design of the percussion input further encouraged this. In order to produce a sound the performer must tap, slap, hit, or otherwise produce a strike on one of the four electrodes. Because the percussive surface doesn't sit on a horizontal surface like the skin of a drum, this presents a range of other interactive options: The instrument might be dropped into the hands, thrown and caught, tossed from hand to hand, held with one hand and tapped with the other, and so on.

The performer also does not have to use their hands. The embedded system recognises a touch when in contact with a conductive surface, which might be any part of the skin.

Again, constraint was applied to the affordances of Keppi. The sound palette was limited, and there was no way to produce sound other than by striking it with a part of the body.

Material considerations

The housing of Keppi is made out of a cardboard poster tube. This presented the advantage of being sturdy yet lightweight, and was a workable size that was not so big that it was hard to grip, or so small that it was fiddly to handle.

The outer materials were selected based on functionality and visual appearance. The electrodes were made with conductive tape which is shiny and reflective, and I wanted to contrast this with a matte material for the spaces between the electrodes that held the lights.

The speakers in the ends of the instrument were held in place by mounts made from the original plastic plugs of the poster tubes and cutting out a space for the speaker in the middle. Using the original plugs meant that the housing was exactly the exact right size. I further reinforced these with an acrylic ring that attached to the speaker and mount with screws, and was covered with plastic mesh to give it a finished look (see Figure 6.11).

The tape that formed the separation between the electrodes, and through which the LEDs shone, was athletic tape for sprained joints. This tape is extremely thin, extremely adhesive, is entirely matte and is textured to the touch, making it hard-wearing, easy to grip, and with a low profile to avoid a ridge on the instrument that might get in the way of playing.

Electronics

The round housing of Keppi required a rethinking of how the internal electronics would be built. (Refer to Appendix H for a detailed diagram and

Figure 6.11: Detail of speaker housing made from the poster tube plug and finished with a face plate.

labelled photograph.) All connections terminated at a breadboard to ensure a firm connection, and cables in turn connected these inputs to a central Bela unit that does all sensor processing and produces sound output.

In order to minimise the number of wires on the inside of the instrument I adopted a cut-and-paste approach to assembly using copper tape. First, I mounted the LEDs in holes drilled in the tube housing, and then soldered the legs to a strip of copper tape (first attaching 220Ω resistors to one leg). Each half of the tube had 3 LEDs of the 6 LED group, and the halves were joined with a jumper wire that went over the hinge.

Each ring of LEDs operated from a single digital pin. Ground was supplied to each strip of LED anodes, and a 5-way cable connected the 5 strips to the breadboard, where they were then attached to 5 of Bela's digital pins.

There was a network of 4 piezos underneath each electrode. These were 15mm piezos soldered together into two networked pairs. The two halves of the piezo networks were connected over the hinge, and a group of 4 collector wires connected all 4 networks to the resistor ladder on the breadboard. which led to a cable that attached to the Bela unit's analogue inputs 0-5.

The capacitive touch was handled using an MPR121 board. The electrodes on the outside of the tube were each made of one large piece of conductive aluminum tape wrapped around the tube and folded over the edge where the instrument opened. This overhang was where the electrodes connected to the MPR121.

In the instruments that had a disfluent behaviour an accelerometer was also used. This was attached to the underside of the breadboard.

6.4.4 Programming and sensor processing

The code for Keppi is publicly available on Github[7].

MOAI's hardware design separated velocity from the sample. The piezo sensed velocity and that was applied to a triggered sample, but the sample was chosen according to the bounce behaviour of the box as sensed by the accelerometer. I iterated again on this separation of velocity and playback when designing Keppi.

This separation was advantageous for two reasons. Firstly, typically both the onset of the strike and its velocity are sensed by a piezo, but Keppi had four piezo networks, one under each electrode, and there was significant cross talk between them when the instrument was struck.

For this reason, I separated sound triggering from the applied velocity. The signal from piezo networks under the electrodes were continuously sampled by the Bela system, but this data was only used if the system detected skin contact on a given electrode via capacitive touch.

Motion was sensed via an accelerometer, which processed the data from the X, Y and Z axes. A 2nd order low pass filter to reject the high frequency motion that would be caused by each strike.

Quantity of motion was used to control the lights. First, the square of the difference in motion for this sample was calculated:

$$motion_{axis} = (previousValue - currentValue)^2 \tag{6.1}$$

Then, quantity of motion was found by determining Euclidan distance:

[7]https://github.com/disastrid/Keppi_2017

$$totalMotion = \sqrt{motion_x^2 + motion_y^2 + motion_z^2}; \qquad (6.2)$$

If quantity of motion decreased over time, the lights would tick down. If it increased, the lights would tick up.

6.4.5 Designing the test condition

For Study 3, Keppi was produced in three versions: No disfluency, mild disfluency, and heightened disfluency. All versions were physically identical, and differed in terms of behaviour.

The disfluent behaviour was as follows: If Keppi was not moved sufficiently, the 5 rows of LEDs would shut off in sequence (a behaviour termed 'tick-down'). When all LEDs had shut off, Keppi ceased to make sound, and had to be moved in order to charge it up again. The difference between mild and heightened disfluency was in the speed in which this tick-down occurred: The version with mildly disfluency had 650ms between ticks, and the version with heightened disfluency had 1500ms between ticks.

The reason for this particular disfluency was two-fold. First, it provided a time-based element that added a sense of urgency — the player could not choose not to engage with the instrument, or it wouldn't make any sound at all. Secondly, the requirement to move the instrument through three dimensional space was not a simple need for interaction, but a need for physical manipulation. I was interested in the additional number of things that could go wrong when a player also had to move through space, and not, for instance, simply interact with the instrument as it lies safely on a table.

This was a direct reaction to the term 'safe', a term that is often used pejoratively to refer to a performance where the performer may be technically perfect but is judged to have not pushed themselves or engaged their skill in a meaningful way. This is a common criticism levied at DMIs, not necessarily because they are 'safe' but a judgement that likely stems from the control dislocation characteristic of a computer instrument — if it doesn't need to be moved because there are no strings to pluck, no membranes to hit, no reeds to vibrate, then it's understandable why physical engagement is not necessarily the norm.

The 'threshold' above which a quantity of motion needed to rise was implemented to ensure that tiny movements would not re-charge the tick down behaviour.

6.4.6 Interaction and sound design

The sound for Keppi was generated with the Chromaphone 2 plugin for Logic[8]. This library allows for percussion synthesis by specifying various real-world parameters (mainly the shape and material of the surface, and various subtle adjustments therein).

In order to connect Keppi's sound to its appearance I synthesised a sound using the parameters for a rigid, hollow, and resonant pipe, adjusting until I felt it matched the physical design. I exported a core tone, and then transposed it to produce the family of four sounds (one per electrode).

6.5 Reflective evaluation

6.5.1 MOAI: Reflections

I consider MOAI to be a successful musical instrument, and I certainly enjoy playing it. The reason for this is that the interactive pluralism makes room for many percussive approaches, the Formalist approach resulted in constraint was effectively applied making the visibility is effective for the audience, and the bounce motion provides intriguing tactile feedback.

The thing that makes me keep coming back to percussion is that there are so many different ways to approach it, and that each percussionist has a style that they have developed through long-term consideration of their body in space, and how it relates to their instrument. I wanted to make MOAI an instrument that retained this. Though it was developed for Ex-Easter Island Head who have a certain existing style (they stand to play, their instruments are on tables, and so on), MOAI does retain the characteristics of drum kits that give rise to a plurality of approaches.

Firstly, though there are three elements connected to one control box, the performer is free to decide how these are arranged. This is similar to a drum kit, where every player has a way they set up that supports their physical characteristics (such as reach), and their style (such as creating room to support two-handed cymbal playing). MOAI is flexible in this way, needing only to be clamped to a table.

[8]https://www.applied-acoustics.com/chromaphone-2/

Secondly, MOAI's main interaction is a strike, and MOAI's form factor, with the elements' large flat top surfaces that are tilted towards the player, support that. The rest is up to the player, as the method of strike is core to the quality and style of percussion practice. Further, there is no limitation on what can be used to strike, and the player can choose their favourite tool or explore others.

Constraint in form factor means that the MOAI elements' only salient feature is their shape, and the motion created when they are played. I took this approach because of the nature of drum kits, which are independent objects until they are activated, connected, and made into a whole by the interaction with a percussionist. I wanted the MOAI elements to be brought to life not by their presence as objects, but by the process of being played.

The materiality was key here, and the matte black, skin-like surface makes them interesting but not particularly remarkable to watch. The physical form factor also makes them unusual, but believable as percussive surfaces — there is no other obvious affordance. I have found that in playing MOAI, and in watching others play it, that the rounded front lends itself to being grabbed, as its shape is amenable to a hand.

I applied extreme constraint to the sound palette. At first I was concerned that it was too limited, but I reasoned that different parts of a drum kit are also extremely limited in their sound character but this does not make them limited in their application (the snare drum, for example, can have variations in dynamics, gesture, rhythm or genre in which it is used, but is always has the sound of a snare drum). I feel this was effective, as it meant MOAI acts like a drum kit and is in that way familiar to percussionists, but also means that the mapping is very limited, supporting visibility for the audience.

The bounce action of MOAI is the most interesting part of the instrument, and the aspect that is not relevant to the results of Study 2 and therefore not discussed, so I will provide some discussion here. The bounce rate provides a novel aspect of the instrument's visibility to the audience, which was effective, but more interestingly this bounce provides a tactile aspect to the performer that I had not expected.

Because each element has about the same mass, they oscillate at a constant rate on the steel slats to which they are attached. When playing a constant rhythm, there is an extremely satisfying physical rhythm that is easy to establish with the box's constant bounce rate. It's easy to get into a constant rhythm, even with eyes closed — there is a 'feel' to this interaction that I have not experienced before with other percussion instruments.

The speed of this bounce rate is adjustable depending on where the slat is clamped to the table. I did some investigation of this and found the BPM of various slat clamp points by analysing the accelerometer data. I am curious to do some testing to see if it would be possible, for example, to play complex polyrhythms by clamping the boxes at different slat lengths for the right and left hand. At any rate, this is an extremely intriguing aspect of MOAI, that was unexpected but one that I intend to develop further in the future.

6.5.2 Keppi: Reflections

I was satisfied with the outcome of Keppi. There were some technical problems during the performance, which are described in Chapter 8, but as an instrument I feel Keppi was successful for the study and remains interesting to play. Similar to MOAI, this success is based on the success of interactive pluralism, effective constraint, and unexpected features that beg further exploration. I will describe each of these in turn.

Interactive pluralism was a design value that again was applied to this design process, and, as was the case with MOAI, I considered carefully what the important aspects of percussion were, and how I could preserve these but bring them into a new context. For Keppi, I wanted to retain this straightforward cause and effect relationship that is characteristic of percussion, but change it slightly so the player had to negotiate a three dimensional object and not a static surface.

The evidence that Keppi achieved interactive pluralism is in the performances of the study. The six performers who participated in Study 3 each developed a diversity of ways of playing the instrument, and their existing styles of playing were evident. Keppi was played standing, sitting, rhythms were added with stomping, it was played with hands, fingers, knees, feet.

Constraint was used in the physical design of Keppi. I didn't want it to reference other percussion instruments but at the same time wanted the interaction to be clear, so I kept the action-to-sound relationship extremely straightforward. The sound palette was very limited which I think supported the clarity of action-to-sound, and the resonant tube sound suited the form factor of the instrument and also was true to the overall functionalist approach (but I would like to explore sounds where the pitch quality is less salient). The materials used — conductive tape and athletic tape — kept the visual information limited but provided clean lines and did not clutter the object with features.

Keppi had some interesting unexpected features. Though I had spent a lot of time playing it during the design process, I had not understood its visual potential until I saw others play it on stage. Keppi's appearance is understated, but quite beautiful, and by limiting the features, paying attention to the surface, and carefully considering the number and placement of the LEDs the result is intriguing but believable as an instrument, as opposed to an experimental device or toy. Keppi's stage presence is more than the sum of a poster tube, athletic tape, and aluminium foil.

Because of the visual aspect of Keppi, and because I have proved that they can be produced in multiples, I am interested in exploring Keppi's ensemble potential. I would very much like to collaborate with other percussionists to compose and perform percussive pieces that would take advantage of the visibility of the tick-down behaviour, and explore a range of musical styles in which to use Keppi.

The challenge that remains is a technical one. Keppi's speakers simply aren't loud enough in their current form, so I need to come up with a way of transmitting the sound output to a PA, or exploring other ways of creating the sound. This is a considerable challenge as part of Keppi's success with players was that it is entirely wireless and self-contained. Additionally, there are some subtle aspects of the capacitive sensing that still need to be resolved to make it reliably responsive. However, when these technical problems are solved in the next iteration, I feel Keppi will have real potential as an instrument for percussion performance.

6.6 Summary

In this chapter I detailed the rationale for producing the instruments for Studies 2 and 3, MOAI and Keppi.

First, I presented the broad issues that governed the production of these instruments. This included a discussion of how, by producing my own instruments, I was able to precisely control the investigative variables which would otherwise have to be compromised on some level if I opted to use an existing instrument. Additionally, I explained my approach to interaction design in this context, and detailed the specific approaches to embedded hardware and sensor design employed in this context.

Secondly, I detailed the specific design and production decisions for each instrument in their own particular context. MOAI, designed for a gesture study, was produced at two scales in order to affect the gestural affordances presented to the player. Keppi, designed for a study examining visible risk, was designed in multiples, each visually identical but differing in their disfluent behaviours.

Finally, I presented reflections on the success of the approach described, including interesting features of the instruments that arose during this process but are not directly relevant to the study outcomes, aspects that need to be improved, and future directions of inquiry. Although these outcomes and reflections do not directly impact the the study results, they still serve as an important documentation of the design process for these two DMIs.

Chapter 7

Study 2: Examining gesture

This chapter describes Study 2, which examined the impact of gesture scale on audience perception of DMI performance, and explored whether gesture scale could be influenced through the physical design of the DMI.

The motivation for this study and the central research questions are specified in Section 7.1. Section 7.2 describes the design of this experiment, including the musicians, the study context, and the method of carrying out this study.

The study results are reported in Section 7.3. These results include statistical analysis of the post-hoc quantitative data, thematic analysis of the qualitative data, statistical analysis and description of the real-time data, as well as reporting of the findings from video analysis. These results are summarised and contextualised in Section 7.5. The results are also described in relation to the study question, the larger research questions guiding this thesis, and the existing literature. Here I also describe the limitations of this study, and indicate how these results will be carried forward into the final experiment.

7.1 Motivation and questions

This study examines the relative effect of gesture on audience perception of DMI performance. Ideas around gesture in the content of DMI research is discussed in depth in Section 2.3.1, and in this section I summarise these ideas as well as present concepts that directly influence this research.

The NIME community is intensely interested in gesture. In 2014 Jensenius found that NIME uses the word ‘gesture’ in an average of 62% of publications per year, far more than other related fields (SMC: 34%; ICMC: 17%) [105].

Gestures are a core component of music, and act as a ‘bridge between movement and meaning’ [82]. Leman [127] has presented theoretical insights on embodied musical cognition, suggesting that music is both *multimodal* (sensed with a combination of auditory, visual and sensory information) and *embodied* (closely linked to bodily experience). A study by Tsay [194] supports this: By listening only to the audio of the competition, amateurs were poor at picking the winners — but experts were similarly bad, despite their musical knowledge and experience. Tsay suggests that this is because the physical and visual qualities of performance are fundamental to the way we form musical judgements, and that without this essential information experts are unable to make discerning judgements of quality.

The sound produced by traditional instruments is tightly bound to the gestures used to play them: The performer gesture generates the energy to produce the sound, thereby determining the amplitude, pitch and dynamic and timbral qualities of the sound [157]. In the case of DMIs, that tight coupling between the action and the sound produced is entirely optional, thanks to the miniaturisation of computers the the labour-saving qualities of electricity [56].

There are many well-established frameworks for understanding that nature of gesture in the context of NIME, such as [74, 154, 43, 209]. There are also studies comparing diverse instruments [138] and the diversity of performance on a single instrument [109]. Wessel and Wright mention that there “should be some sort of correspondence between the ‘size’ [of the gesture] and the acoustic result” [209], but they do not offer any specifics on gesture size, or how the instrument itself might affect this relationship. In order to address this gap in specific gesture knowledge, this study examines how the size of instrumental gesture changes audience perception by comparing audience feedback on one instrument made in two sizes, and considers how the physical design of a DMI impacts the gestures used to play it.

7.1.1 Defining ‘gesture’

Cadoz and Wanderley, in an early interdisciplinary study of gesture, state that gesture does not have one common definition [40]. NIME has borrowed heavily from HCI ideas about gesture, from adopting broad definitions such as 'a motion that contains information' ([123], quoted in [26]), to applying Fitts law to musical gestures [204], to developing systems to understand the affective content of gestures [42]. Because of this wide investigation into gesture and a wide array of overlapping definitions, it is important to precisely define our use of 'gesture' in this context.

In the context of this study, I consider the notion of 'gesture' within Miranda and Wanderley's 2006 concept of *instrumental gesture*: '[I]t is applied to a concrete (material) object with which there is physical interaction, and specific (physical) phenomena are produced during a physical interaction whose forms and dynamics can be mastered by the subject.' [141] In other words, in this context gesture is defined as the musician's physical action applied to an object that produces sound.

7.1.2 Gestural affordances of the DMI

Leman [127] presents a theory of the body building up gesture 'repertoires' as it mediates between the physical world and subjective experience, made up of gesture/action consequences. Though there is a huge range of performer decision, history, and knowledge that will determine their exact method of playing (as established by Jorda [109]), the physical design of the DMI impacts this gesture repertoire by presenting certain *affordances*. Affordances are defined by Gibson [81] as what an environment offers to an animal (or human actor) within it. This relationship can be described as a mapping between environment's properties, and the actor's potential actions. The affordances of an instrument have been used to study the microdiversity of playing styles [214], but there is yet no specific use of affordances to adjust gestural repertoires.

There is, however, evidence that affordances of a DMI does affect performer decisions during play. Jack et al. [103] makes the argument that the physical design of a DMI can influence a performer's gestural language used to play it. This is an extension of Tuuri et al.'s theory of 'push' and 'pull' effects [196], which suggests that though humans control interfaces, these interfaces also affect the choices we make in that control through their design.

7.1.3 Gesture and multimodality

Music is a multimodal experience, meaning visual and other sensory factors profoundly influence the way audiences perceive it. There is general consensus that gesture is important: Fyans et al. [71] suggest that audiences must have a way of relating gestures to sound, to build a mental model of how the instrument and understand the performer's intent; Schloss [175] states that since the gesture and sound relationship is not necessarily one-to-one, we must understand what works in order to communicate to audiences; Fels et al. [68] situate the importance of this communication around the notion of transparency, asserting that the audience and performer must share a common understanding of input-output mapping in order for communication to take place; Sheridan [180] argues that an audience must understand the *performance frame* created and used by performers.

Despite this consensus, there is not much indication of how a gesture to express a given message might look, or how we might craft interfaces to support effective gestures. Schloss [175] states that a 'visual component is essential to the audience' and that 'effort is important', but goes no further.

There is some indication of effective gesture characteristics by Reeves et al. [169] in their consideration of gesture in the context of 'performing' with a computer. They define the input/output relationship to be made up of *manipulations* and *effects*. By placing these two elements on axes from *hidden* to *amplified*, and a taxonomy of gesture can be extracted (illustrated in Figure 7.1):

magical: (amplified effects, hidden manipulations)

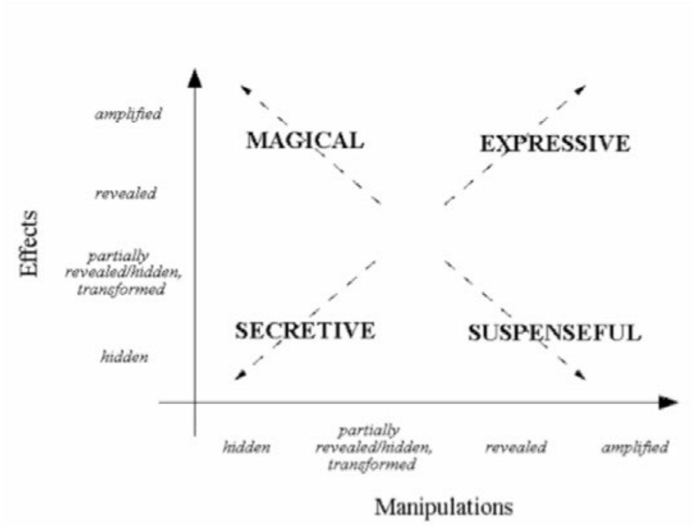
secretive: (hidden effects, hidden manipulations)

suspenseful: (hidden effects, amplified manipulations)

expressive: (amplified effects, amplified manipulations)

Considering expressivity as a goal of music performance, this taxonomy suggests that both the performer's manipulations and the effects should be amplified in order to achieve this. However, it is missing considerable detail: This taxonomy does not tell us, for example, how big a gesture has to be to achieve amplification. Further, there is no indication if this amplification is a function of the gesture itself, or the scale of that gesture. There is also no

Figure 7.1: Illustration of Reeves et al.'s taxonomy of gesture [169].



discussion of how the physical design of the DMI might impact performer gesture and, by extension, support its amplification.

I was motivated, therefore, to test whether the size of the gestures used to play an instrument have any effect on audience perception. Particularly based on contributions by Jack et al. [103] and Tuuri et al. [196], there is precedent to suggest that a difference in size may indeed affect the scale of gesture, and I wanted to create a method of direct comparison using one instrument. By using a custom-designed DMI in two versions, one large and one much smaller, a participating musician was presented with one set of sound controls and one set of affordances, but at different scales. In this way, the DMI's physical design had direct impact on the scale of the gesture, presenting the potential for insight on how the scale of a given gesture might affect audience perception - and potentially shedding light on what an effective gesture might look like.

7.1.4 Research questions in this study

The specific question of this study is as follows:

What is the relative effect of gesture size on audience perception of DMI performance?

As well as this primary research question, this study contributes to the following aspects of the overarching questions of this research:

RQ3[1]: The DMI designed for the purposes of this experiment was MOAI (described at length in Chapter 6). This instrument was produced in two verisons identical in design, form factor, interaction and response but differed in size. For this study, the performer played the same piece on both versions, meaning that there was control of the factors facing the performance but there was a difference in the scale of gesture.

RQ4[2]: As with Study 1, the button taps from each participant were correlated with their ratings of Enjoyment, in order to determine if a binary relationship exists between 'enjoyment' and 'error'. Further, the real-time data was synced with the performance video footage in order to extract performance features at times of audience agreement about 'enjoyment' and 'error' events.

7.2 Study design

7.2.1 The instruments and musicians

The instrument designed for this study is called MOAI. An in-depth exploration of the design approach, process, sound design, material considerations, and programming of MOAI can be found in Chapter 6.

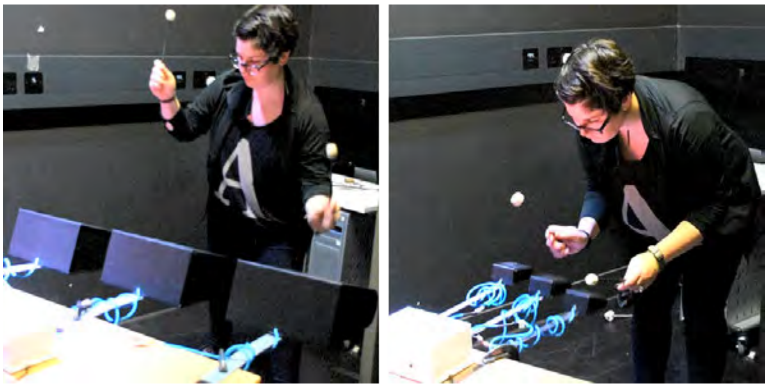
Two musicians were recruited for this study. One was a laptop performer, Joanne Armitage[3]. The other was a percussionist, Ben Duvall of Ex Easter Island Head[4].

The instrument designed and produced specifically for this study, MOAI, was played by Ben. It was produced in two versions: One was large, with each of MOAI's three elements measuring 40x30x20cm, and the other was small, with each element measuring 12x10x5cm. (See Figure 7.2 for an illustration, and refer to Chapter 6 for additional details.)

Ben was recruited for this study because he has a long history of percussion performance with innovative instruments. The initial designs were done with his practice and sound in mind, and he and his band gave input into the instrument's form factor and interaction midway through the design process.

[1]RQ3: Can the physical design of a DMI affect the performative outcomes?
[2]RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?
[3]https://joannnne.github.io
[4]http://exeasterislandhead.com

Figure 7.2: Photographs of me playing MOAI, demonstrating the difference in physical scale of the large and small versions in action. Left: Large version. Right: Small version.



MOAI was designed to be an instrument that was imbued with movement. It was designed to be played with mallets (Ben's preferred method), and the oscillation of the percussion elements means that the instrument moved and undulated in response to being played. In this way, playing MOAI required using gestures that were demonstrative of hitting percussive surfaces.

7.2.2 Precedent effect

Precedent effect, where the order of performances influences the audience experience, is a major consideration in any study of this nature, but was especially important in this instance. With one instrument in two versions, it would be impossible to separate the effects of precedent from this study, as seeing the instrument once would potentially allow the audience to establish expectations of the instrument that would carry on to the following performance.

For this reason, the audience was divided into two, and each audience saw a performance on one version of the MOAI, as well as a laptop performance. Audience 1 saw the performance on the large MOAI and then the laptop, and Audience 2 saw the performances on the small MOAI and then the laptop. The laptop, as it is an instrument that has a gesture set that is extremely subtle, served as a 'control' performance against which to compare the MOAI ratings. This is to control for variations in the two audiences; as the audiences were divided based on where they chose to sit, it was possible that one audience may differ drastically from the other in terms of factors such as personal taste.

7.2.3 Navigating complexity

As discussed in Section 3.3, the issue of complexity in these studies was encouraged where it was musically useful, and controlled where it presented a confounding factor.

Complexity was encouraged in some aspects, as it added to the multimodal experience of live music, and helped to deliver a credible experience. These were:

Recruiting musicians. Recruiting musicians for the study meant that the performances would encapsulate musician's style and creativity. It may have been more straightforward to ask the performers to rehearse pre-composed performances, or to operate within an extremely limited rhythmic repertoire, but this would also remove the opportunity for the audience to watch musicians perform their own music. There is complexity in the inherent variation between two creative practices, but this is more akin to how live DMI performance is typically experienced.

Using a concert setting. I again used the Film and Drama Studio as the location for this concert, for the main purpose of preserving the multimodal aspects of a live concert (instead of, for example, holding this study in a lab setting). Further, this meant the setting, acoustic, contextual and visual aspects of the concert would have a baseline of consistency with the previous study, which would be useful when ultimately combining results for the larger study questions.

As well as encouraging complexity to support the live and multimodal aspects of this study, there were ways in which complexity was controlled in order to produce results. These were:

MOAI and laptop performances used the same sounds. In order to control for timbral differences, Joanne was given the sound set that MOAI used in order to create her performances. This means that the performances were similar in their sounds.

Both pieces were percussive. Although a laptop is not necessarily a percussive interface, Joanne was asked to approach the piece from a percussion perspective. This meant that there would be rhythm present. This was because of the findings of Study 1 (described in Chapter 5), which found that experimental music tended to be less enjoyed than music that aligned with conventions such as regular rhythm.

A piece was composed for MOAI. Instead of improvising on the instrument, Ben composed and rehearsed the same piece on both versions of MOAI. This meant that the gestures were well rehearsed, and that differences in audience ratings could be directly compared.

No projected code for the laptop performance. As this was a comparative study of gesture, the laptop performance had no projected code (though this is conventional for laptop performances). This was to ensure that the audience was judging based on sound and performance, and confounding visual content was removed.

Audience snacks. The audience was divided into two groups, and the Audience 1 watched their performances while Audience 2 enjoyed snacks in the hallway, and vice versa. In order to control for the impact of snacks on musical perception, snacks were served to all before Audience 1's performance in the ten minutes between set changes. In this way, no group was denied snacks or a break before seeing their performances.

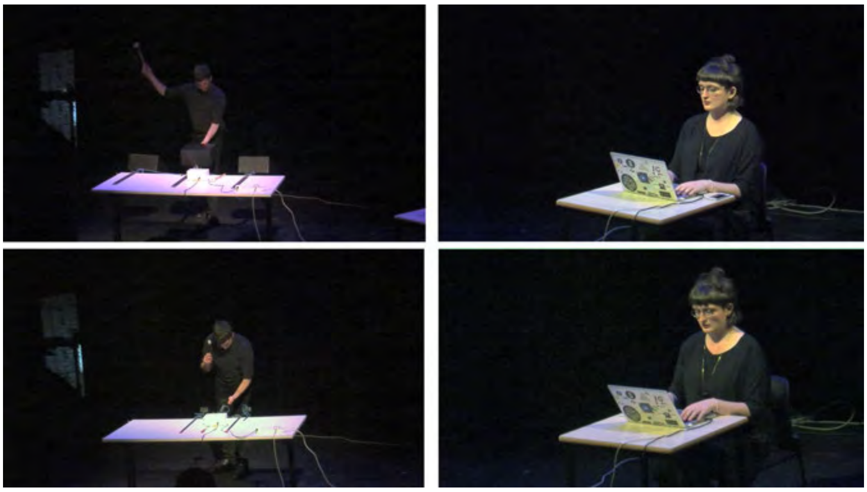
7.2.4 Study method

This study took place on November 10 2016, in the context of an evening concert, on the theme of experimental percussion. Act 1 was an opening performer (Enrico Bertelli[5] doing body percussion), Act 2 was the audience study, and Acts 3 and 4 were solo performances by Joanne Armitage and Ben Duvall (see Figure 7.3 for images of Ben and Joanne).

After Act 1 the audience was divided into two based on where they were sitting, into Audience 1 (N=14) and Audience 2 (N=13). The study was introduced, and the audience was invited to participate. The survey booklets and pens were passed out, containing two post-performance surveys and a

[5] `http://www.enricobertelli.co.uk/`

Figure 7.3: The study performers. Left column, MOAI performances: Top, large MOAI; bottom, small MOAI. Right column, laptop performances: Top, performance 1, bottom, performance 2. NOTE: Stills taken at congruent points in the performance to demonstrate gesture scale.



longer post-concert survey (surveys are available in Appendix I). The scale variable (large MOAI and small MOAI) was not explained to the audience, so they were not aware of how the two sets of performances differed.

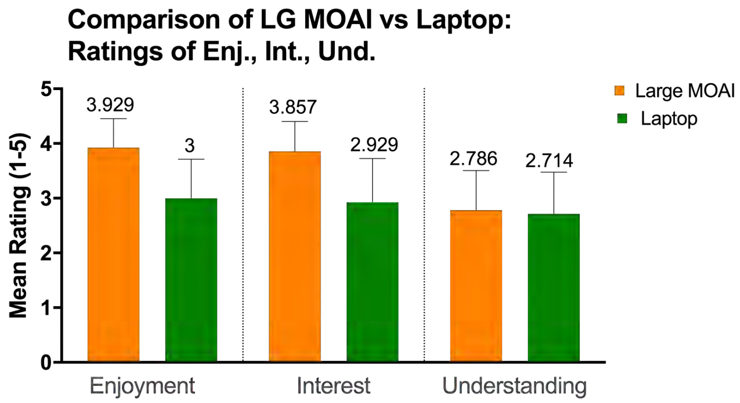
Metrix was introduced to the whole group, and the onboarding video was shown. I then addressed audience questions.

Both audiences went into the hallway to enjoy snacks as the MOAI and laptop were set up (these were hidden beforehand). Audience 1 then came in, watched the large MOAI performance while using Metrix, and then filled out their post-performance survey. They then watched the laptop performance, again while using Metrix, and then filled out the second post-performance survey. Finally, they filled out the longer post-concert survey.

Audience 1 then went into the hallway and ate snacks. The MOAI setup was changed, and Audience 2 entered. They watched the performances while using Metrix and filled out their questionnaires according to the method above.

Then, the whole audience recovened in the performance space, and Joanne and Ben both performed solo pieces of their own.

Figure 7.4: Graphs of mean ratings of Enjoyment, Interest and Understanding for Audience 1, Large MOAI vs laptop. Error bars indicate 95% CI.



7.3 Results: Post-hoc

Post-hoc data was collected from both audiences using written questionnaires (see Appendix I).

7.3.1 Quantitative

The quantitative data is composed of audience rankings of each performance for Enjoyment, Interest and Understanding from 1 (least) to 5 (most), as well as rank ordering of the two performances in terms of preference.

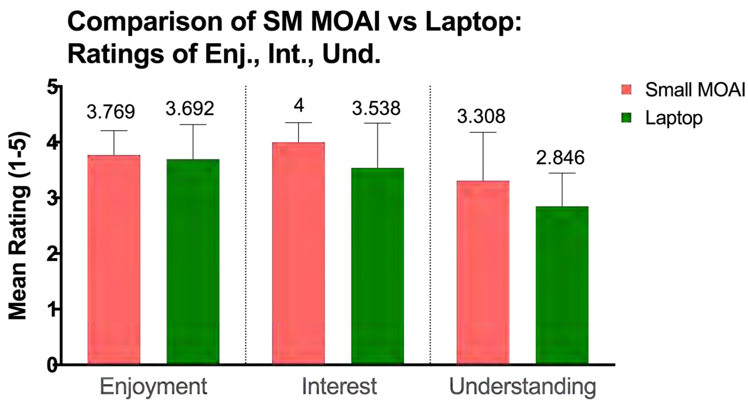


Figure 7.5: Mean ratings of Enjoyment, Interest and Understanding for Audience 2, Small MOAI vs laptop. Error bars indicate 95% CI.

Audience 1: Ratings of large MOAI vs laptop

To determine if ratings of Enjoyment, Interest and Understanding by Audience 1 for the large MOAI and the laptop were statistically significantly different, a two-tailed paired t test was used to analyse the ratings of each aspect.

The ratings of Enjoyment and Interest were statistically significantly different for ratings of Enjoyment (t=2.616, df=13, p=.0213) and Interest (t=2.329, df=13, p=.0366). There was no difference of statistical significance found between the ratings of Understanding for the large MOAI and the laptop performance. (See Figure 7.4 left for an illustration of these results.)

Audience 2: Ratings of small MOAI vs laptop

To determine if ratings of Enjoyment, Interest and Understanding for the small MOAI and the laptop were statistically significantly different, a two-tailed paired t test was used to analyse the ratings of each aspect. For each of the aspects of Enjoyment, Interest and Understanding, no statistically significant differences were found between the ratings of that aspect for the MOAI and the ratings for the laptop. (See Figure 7.5 right for an illustration of these results.)

Rank ordering

In their post-concert survey, both audiences ranked the two performances in order of preference (1 = most preferred, 2 = least preferred).

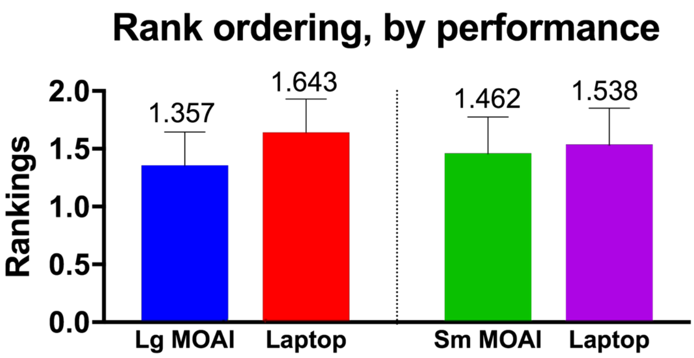


Figure 7.6: Means of rank ordering, Audience 1 vs Audience 2. No statistically significant differences were found. Error bars indicate 95% CI.

A Mann-Whitney test was performed on the rankings to determine if either of the audience ratings differed significantly. However, no differences of statistical significance were found. (The mean ratings are illustrated in Figure 7.6.)

7.3.2 Qualitative

Qualitative data was also collected via the questionnaires. In each post-performance survey three questions were asked:

1. What did you like about the performance?
2. What did you dislike about the performance?
3. How would you describe the performance to a friend?

The data corpus assembled for analysis is composed of the written answers to questions 1 and 2, for each of the performances.

A thematic analysis was performed on this data, using the method described in Section 5.4. The coding and themes generated by this data corpus are available in Appendix J.

The themes that resulted from this process were similar to those from Study 1 (described in Chapter 5. These were:

Sound output: Aspects of the sound, such as the quality of the sound and compositional aspects (Quote: *'The sound of percussion, rhythm'*)

Way of playing: Descriptions that related to the performer, or the way the music was performed (Quote: *'Percussionist is clearly a skilled performer.'*)

Instrument: Statements that related to the instrument (Quote: *'The instrument is interesting.'*)

Experience: Responses that described aspects such as value judgements and emotional response (Quote: *'It feels like the sound is reaching my mind.'*)

There were also a number of responses that were non-responses, such as 'Nothing' or 'N/A'. These were disregarded.

Due to the audience being divided to mitigate precedent effect, the sub-audiences were small in number (Audience 1, N=14; Audience 2, N=13). Therefore, percentage numbers of theme occurrence are not as useful as they were in Study 1 where the audience was much larger.

However, some notable features did emerge when viewing the data corpus in terms of themes. These were:

Mentions of physicality

The large MOAI had several 'like' comments related to the way it was played ('Physicality, rhythm and clarity of action', 'Immediacy'). The small MOAI had one such 'like' comment ('The simplicity of actions') nor either of the laptop performances had 'like' comments relating specifically to the physicality.

Experience statements

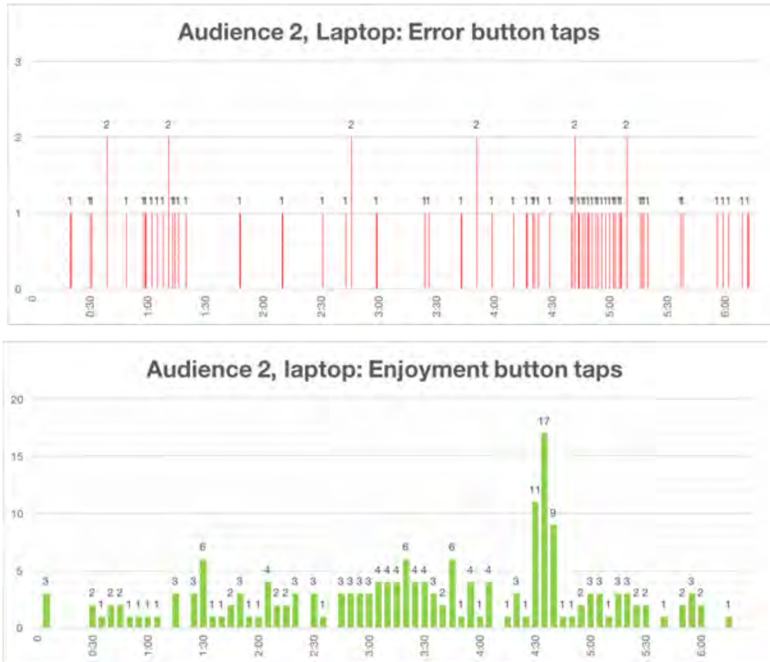
Comments related to the theme of Experience were predominant in the 'dislike' answers that Audience 2 cited for the small MOAI. For Audience 1 they Experience mentions in the 'dislikes' were still present for both the large MOAI and the laptop, but these comments for the MOAI centred on experiences (*'Long time of non-changing drum beats'*) while the comments for the laptop tended to contain on *value judgements* in these experience statements (*'I find live coding music to be tedious to watch'*).

7.4 Results: Real-time

The real-time data was, as in Study 1, collected as time stamps in Unix epoch time with millisecond resolution.

Though there were no browser-based truncations of the data to the nearest second, millisecond-resolution data was too granular for results to be observed, and all tap events were rounded to the nearest second. The 1s resolution again proved too granular to observe 'enjoyment' events, so these were divided into 5s time bins.

Figure 7.7: Visualisation of ‘error’ vs ‘enjoyment’ taps the laptop performance evaluated by Audience 2. All histograms are available in Appendix K.



7.4.1 Observations of the real-time data

As observed in Study 1, the ‘error’ and ‘enjoyment’ data had different characteristics when visualised. ‘Error’ events appeared as spikes in the data, and ‘enjoyment’ events appeared to have a Gaussian distribution, with agreements in time bins tending to build to a peak and tapering off again. (See an example in Figure 7.7).

7.4.2 Comparison of button tap rates

For investigation of button tap rates, all tap numbers were normalised to taps-per-minute to account for the slight variation in performance lengths.

First, I wanted to determine if there was any difference in the tap rate of the buttons between the MOAI performances and the laptop performances.

A Mann-Whitney test was used to compare the tap rate of the ‘error’ buttons between the MOAI and laptop performances. No statistically significant differences were found.

A Mann-Whitney test was also used to compare the tap rate of the 'enjoyment' buttons between the MOAI and laptop performances. Again, no statistically significant differences were found.

This data was further divided by audience. Since these ratings were paired and was nonparametric (according to a Shapiro-Wilks normality test), a Wicoxon signed-rank test was used for each comparison.

For Audience 1, a Wilcoxon test on the median taps-per-minute of the 'enjoyment' button determined that there was no statistically significant difference for the large MOAI and the laptop performance. A Wilcoxon test for the same audience that compared the tap rate of the 'error' button also determined that there was no statistically significant difference for the large MOAI and laptop performances.

For Audience 2, a Wilcoxon test on the median taps-per-minute of the 'enjoyment' button found no statistically significant difference for the small MOAI and the laptop performance. However, a Wilcoxon test for the same audience that compared the median taps-per-minute of the 'error' button did find that the per-minute tap rates for the small MOAI performance (Mdn=0.3085, mean=0.5656) were statistically significantly lower than that of the laptop (Mdn=0.6434, mean=1.144) (p=0.0234).

7.4.3 Correlation of button tap rates with ratings of Enjoyment

A Spearman (nonparametric) correlation was made between the per-minute rate of button taps made by an audience member, and their subsequent rating of Enjoyment for that performance. The intuition was that if real-time and post-hoc data were related and if error and enjoyment indeed had a binary relationship, then there would be a negative correlation between the number of 'error' taps and ratings of enjoyment, and a positive correlation between ratings of enjoyment and the number of 'enjoyment' button taps.

Table 7.1 summarises these results. For this correlation, only users who had an associated data set and some number of taps were included. Users who did not touch the tested button during a performance were excluded — this is because there is no way to be sure that they didn't tap the button because they truly were or were not enjoying the performance, or if they had simply stopped using the system.

Only one correlation of statistical significance was found, between the rate of 'enjoyment' button taps and the post-hoc Enjoyment ratings for the large MOAI performance.

7.4.4 Video data

All performances were recorded on video. The real-time data was synced with the video footage via an audible click in the audio that was made when Metrix was made active. 'Enjoyment' and 'error' events were identified using the procedure outlined below. The video was reviewed at these points in order to extract features of 'enjoyment' and 'error' as identified by this audience.

Identifying events

I used the following rules to identify 'enjoyment' and 'error' events in the video documentation.

'Enjoyment' events: The threshold for an 'enjoyment' event was a consensus of 6 audience members in a given 5s time bin. The enjoyment event was considered to have ended on the next bin that drops below this threshold.

'Error' events: The threshold for 'error' events was a consensus of 2 or more audience members in a given time bin. As there were fewer audience members, I also considered 4 indications in 3 contiguous seconds to also be a point of interest to investigate.

Histograms of the real-time data with the marked events are included in Appendix K.

	Audience 1				Audience 2			
	MOAI (lg)		Laptop		MOAI (sm)		Laptop	
	Error taps	Enjoyment taps	Error taps	Enjoyment taps	Error taps	Enjoyment taps	Error taps	Enjoyment taps
r_s	0.011	0.579	-0.331	0.441	-0.332	0.010	-0.602	-0.088
p-value	0.977	0.041	0.266	0.131	0.318	0.978	0.095	0.824

Table 7.1: Chart of Spearman (nonparametric) correlation of button taps and Enjoyment ratings for each performance. Significant p values highlighted.

Coding the video

Links to the video and coding documentation can be found in Appendix L.

I coded the video by reviewing the video documentation at times of 'enjoyment' and 'error' events and coding what was happening at that time. To code the video, I used the procedure described in Section 5.5.7.

Themes of 'enjoyment' events The codes from the 'enjoyment' events clustered around the following themes:

1. **Consistency:** Established rhythm, consistency, flow.
2. **Novelty/change:** Change, adding of new element or sound.
3. **Way of playing:** Confidence, performer action, control.

Because this was a percussion-focused concert and there was a limited sound palette, it was expected that the comments did not include elements such as pitch and melody as they did in Study 1.

There were more 'enjoyment' events associated with the large and small MOAI performances (9 events for each) than with the either of the laptop performances (2 for laptop 1, and 4 for laptop 2). Themes 1 and 2 were observed in both the MOAI and laptop performances, but theme 3, Way of playing, was associated only with the large MOAI performances. This suggests that the larger interface was more visibly controlled by the performer, and this was associated with audience enjoyment.

Themes of 'error' events The codes from the 'error' events were clustered around the following themes:

1. **Pattern related:** Inconsistency of rhythm, interruption
2. **Unexpected elements:** Sudden sound, unexpected start/stop, change
3. **Performer actions:** Intention, way of playing
4. **Trivial technical errors**

There were more error events associated with the first and second laptop performances (Laptop 1, N=8; Laptop 2, N=12) than there were for the

MOAI performances (Lg MOAI, N=2; Sm MOAI, N=4). For the small MOAI performance's 4 error events, these appeared in the midst of rhythmic flow and it was not clear what the error event was referring to, so only 1 error event was coded, and it was a pattern that had gone on for a long time and it seemed unclear that this was intentional.

The large MOAI performance's error codes were related to themes 1 (inconsistency of rhythm) and 4 (technical error, as a box was accidentally muted).

The laptop errors were spread across all 4 themes, but were largely related to theme 3, or performer intention. There were variances and breaks in the rhythm, but it was unclear whether this was intentional or a mistake.

7.5 Discussion

7.5.1 The central question

The central question of this study was:

> What is the relative effect of gesture size on audience perception of DMI performance?

Size (probably) matters

The large version of MOAI was rated significantly higher than the laptop for Enjoyment and Interest. While the small MOAI also rated higher than the laptop for these qualities, the difference was not statistically significant. While not conclusive, this suggests that the effect could be because the larger MOAI produced a more visible interaction, resulting in what Reeves et al. [169] would term 'amplification'.

The qualitative data was the source of more insight. Experience-related comments predominated the cited dislikes for both laptop performances, and this was especially true of Audience 1's comments. This could indicate two things. Firstly, that dislikes related to the theme of Experience result in lower ratings. Secondly, this could also indicate the establishment of expectation; it is possible that seeing the large MOAI and its large gestures established an expectation that was not met at all by the extremely low-gesture laptop performance. The results of this study are not conclusive enough to point to either of these, or make further suggestions, beyond simply that the larger MOAI was more enjoyed.

When comparing the qualitative data between each MAOI performance and its paired laptop performance there was no obvious trend of distaste, and negative words such as 'boring' were equal between them. However, comments related to the large MOAI cited terms such as 'immediate' and 'physical', descriptors that were not applied to the small MOAI at all.

This suggests that the larger MOAI had more visible gestures, and that the cause and effect was more obvious. This makes a case for visibility in a DMI, and that control dislocation may continue to be a confounding factor for audiences, even when there is a physical connection — the physical connection may have to be at a large enough scale to be salient and, therefore, more effective.

This ambiguity of scale, however, highlights an open question: How can we measure gesture, to arrive at more detailed insights? Motion capture, for instance, may be an option, as this would allow for modelling of the performer gesture. But, motion capture has to take place in a specialised environment, and with the concentration on preserving the concert setting throughout this research that is not possible, though it may be appropriate for other studies. Another interesting option would be to do some analysis of the performance footage to get more granular data on the relative bigness or smallness of gesture, or to learn about the range over time. Clearly, this is an intriguing area for further study.

The most compelling data related to visibility came from the video coding. The audience indicated 'error' events most often that were related to unclear performer intention, and 'enjoyment' events, particularly for the large MOAI performance, related to aspects of performance (such as flow and performer-related events, such as purposeful and skilful strikes).

This suggests that the audience responded to *visible skill*. Although Joanne is an extremely skilled laptop performer, her skill is not visible through her gestures while performing. Further, when she uses breaks in rhythm, it appears that the audience, not knowing if that break was intentional, considers this to be an error. Contrastingly, Ben's skill as a percussionist is apparent when he plays MOAI, in the confidence of his strikes and the purpose of his movements.

More study is needed to make conclusive statements about the role of gesture size and the audience perception of skill, but these preliminary findings do align with the findings of Fyans et al. [72] on perceptions of skill in DMI performance. They found that control and effort were perceived by audience members as markers of skill, and that skill is 'an embodied phenomenon'.

Gurevich and Fyans [86] suggest that the listener is sensitive to the intentions of the performer. If this is taken to mean that understanding intention is important to audiences, the above findings reinforce this. It is also notable that this study was done using video recording and the data collected was via structured interviews, meaning that the results of this study add vital real-time insight from a live context.

Both these works raise the concept of *instrumentality* as defined by Cadoz [39] (as briefly discussed in Chapter 2). Though the term DMI includes a huge range of music-making interfaces, methods and processes (as reviewed in Chapter 2), Cadoz proposes that the definition of 'instrument' has been overstretched when applied to DMIs, and that 'instrumental interaction' should be limited to mechanical processes through which there is energetic exchange. In this way, the laptop — while certainly a DMI — does not have the same instrumental quality as the MOAI, which encourages percussive gestures that excite a surface in order to create sound.

It is notable that, in this study, the larger version was rated higher for Enjoyment by the audience, and the audience also appeared to respond in real-time to the embodied aspects that relate to player skill. This suggests that while both MOAI versions are played with instrumental interactions, a larger size supports the perception of skill, possibly because that instrumental interaction is made more visible. Because this larger size also rated significantly higher, it is possible to conclude, by extension, that perceptible skill is important to audiences. This study cannot conclusively suggest this, but it does indicate an route of enquiry that could lead to useful insights and firmer conclusions.

The video coding also brought up the concept of intent. 12 of the 18 'enjoyment' events for both versions of MOAI pertained to performer action, whereas only 1 of the laptop 'enjoyment' events pertained to performer action. Conversely, the laptop 'error' events were largely centred around unclear performer intention. This suggests that because there were no gestures associated with the sound output it was difficult for the audience to tell if rhythmic inconsistencies were stylistic and intentional, or unintentional mistakes.

7.5.2 Revisiting the larger research questions

As well as the central question above, this study also informed two of the larger questions guiding this thesis:

RQ3: Can the physical design of a DMI affect the performative outcomes?

RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

I will consider the results relating to each of these in turn.

The influence of the physical design

The 'outcomes' to which this question refers are from the previous question, RQ2, which asks: 'What is the impact of visible risk on audience perception of DMI performance?'

If we consider a visible risk a state where skill becomes apparent, then the findings suggest that the MOAI at the larger scale did affect this. Given that the composition played on MOAI was the same for both versions, it also suggests that the musical outcomes were more enjoyable for one version than the other. Again, the data from this study is not conclusive enough to make firm conclusions, but the results do indicate that an aspect of the physical design — scale — had an observable effect.

Combining post-hoc and real-time data

This study was the second time the combined methodology was carried out and the two data sets compared, and I consider this study a second iteration of the design of this methodology.

The same methods of analysis were used (correlating ratings with enjoyment and error taps, comparing tap rates, identifying events in the video, coding the video and analysing these codes thematically). Some of these were fruitful, such as the insight on skill that emerged from the video documentation.

The correlation of the real-time data with the Enjoyment ratings found one significant correlation, but this is not conclusive enough to draw any firm conclusions. Though other studies have been carried out in this space with audiences of a similar size (notably the study by Gurevich and Fyans which I cited above [72, 86]), it is possible that this combined methodology is only useful for audiences over a certain size.

That said, reviewing the video documentation at the times of 'enjoyment' and 'error' events was insightful, as it added the aspects of skill discussed above, and intention (which I discuss below). These are useful observations that add to the findings of Study 1 about the nature of 'enjoyment' and 'error' as perceived by audiences. It is also important to note that these insights were not observable in the post-hoc qualitative data, so the combined methodology, though it did not always produce statistical results, did successfully find indications that would have otherwise remained unexplored.

7.5.3 Limitations

This study also has limitations, which should inform the interpretation of its results.

Like Study 1, this is a study that was limited to two performers, two instruments, and two audiences. This is not a recommendation on a particular dimension for DMIs — it is possible that the large MOAI was still not large enough to elicit effectively 'amplified' gestures, and only further targeted study can determine this.

It is also important to note that comparing large and small MOAI cannot be separated from the quality of the laptop performances: The large MOAI performance receiving significantly higher ratings than the laptop performance may simply indicate that the second laptop performance was better than the first. Though the ratings of the laptop between the two audiences were not statistically significantly different — indicating that the laptop performances were similar — this is still a confounding factor that bears consideration.

7.6 Summary

This chapter described the second study in this thesis. This study, taking place in the context of an evening concert, examined the relative effects of gesture size on audience perception of DMI performance, in order to gain insight into how a gesture might be considered, according to Reeves et al. [169], to be 'amplified'.

Using MOAI, a DMI designed and built for this study in both a large and a small version, this study investigated the following question: *What is the relative effect of gesture size on audience perception of DMI performance?*

In the context of an evening concert, two separate audiences watched a pair of performances. Audience 1 watched a performance on the large version of the MOAI, and then a performance on a laptop. Following this, Audience 2 watched a performance on the small version of MOAI, and also a performance on a laptop. Both performances were played by the same musicians (Ben Duvall on MOAI and Joanne Armitage on laptop). The audiences were entirely separate because it was impossible to adequately control for precedent effect. For this reason, the laptop performance was used as a control for comparison, as comparing the unpaired opinions of two audiences on the different versions of MOAI would offer little insight.

The results of this study suggest that the size of the DMI likely adds to the visibility of gesture, and that in this way audiences are able to perceive performer skill. Skill, suggest Fyans and Gurevich [72], is an embodied phenomenon that is perceived by control and effort, and therefore MOAI, given its instrumental nature (as defined by Cadoz [39]), resulted in perceptions of skill. Considering that the large MOAI was rated higher, this suggests that the scale of a DMI has a positive impact. Although the results of this study are not conclusive, this does highlight an avenue of future study that could potentially yield valuable insight.

This was also the second application of the combined methodology. Coding the video at the times of 'enjoyment' and 'error' events and grouping these codes into themes did reveal important insights about skill and intention (though part of this lack of results was that 'error' and 'enjoyment' still appear to be unrelated). However, analysing the real-time data and the post-hoc ratings did not give rise to firm conclusions. This may be due to the smaller number of audience participants in this study, as the audience was subdivided into two groups to address precedent effect. Despite this reduced effectiveness, the real-time data did yield important insights when applied to the video documentation.

7.6.1 Carry forward

Three important insights emerged from this study. First, the design and production of a DMI for experimental purposes was successful, and allowed the study of precise questions. Further, creating multiple versions of one DMI also allows for a direct axis of comparison, which was insightful.

Secondly, the emergence of skill in this study lends dimension to the findings from Study 1. Because skill seems to play an important role but the understanding around it is quite murky, I was interested in bringing this into the third study.

Finally, the finding that the combined methodology is less effective with fewer participants was a good lesson. Since subdividing the audience reduced the numbers of audience participants for each set of performances to levels that did not yield many conclusive statistical results, I wanted to ensure that the final study would not have any subdivision or require separation of audiences, in order to increase the possibility of more conclusive outcomes.

Chapter 8

Study 3: Exploring disfluency

This chapter describes the final study of this thesis, which focuses on the impact of visible risk on audience perception. Specifically, this study examines the relative effects risk produced through disfluent instrument behaviour.

Disfluency is defined as the 'metacognitive experience of difficulty associated with a cognitive task' [60]). Making something intentionally challenging to process (for example, by printing it in a font that is hard to read), has been shown to result in heightened cognitive processing. In Section 8.1 I explore some of the existing research around disfluency, explore how this played a role in Keith Jarrett's performance of *The Köln Concert*, and describe the central question of this study.

I then detail the design decisions made for this study in Section 8.2. This includes the design of the instrument and the selection of performers, as well as the context of the study and the method used in carrying it out.

In Section 8.3 I report the results of the qualitative, quantitative and real-time data gathered, and finally in Section 8.5 I discuss how the results provide insight into to the overarching question of error in this work. I also discuss the role of visible risk, and the role of visible skill, and also indicate implications for DMI designers.

8.1 Motivations and related work

In this final study, I explored the effects of a risk state, produced by disfluent behaviour in a DMI, on audience perception of DMI performance. The instrument used for this study was Keppi, the design of which is described in detail in Chapter 6. This section examines what disfluency is, and why it motivated this final study. Finally, I discuss how this notion intersects with existing knowledge in the DMI literature, and how this helped to shape this study's central question.

8.1.1 Disfluency and its potential for DMI design

The usefulness of disfluency, defined as 'the experience of processing difficulty' [192], has been demonstrated in cognitive science research. Alter et al. published the results of a study [5] in which they tested the influence of disfluency in text on the process of mental reasoning, incorporating disfluency by printing the content in a difficult-to-read font.

The result, they suggest, is that disfluency produced heightened cognitive function, and therefore better test outcomes (a conclusion that is counter-intuitive to both students and educators [60]). The authors suggest that 'experiences of difficulty or disfluency appear to serve as an alarm that activates analytic forms of reasoning that assess and sometimes correct the output of more intuitive forms of reasoning.' In other words, processing difficulties force the test taker to slow down, consider the question, and reason using their cognitive skill, and not just rely on their intuition or first impressions, which are often unreliable.

Cognitive tests are not creative acts, but these findings hint at applications for disfluency in cognitive tasks that have a temporal element, which certainly includes DMI performance. I was intrigued by the suggestion that it is not a lack of fluency that poses a challenge, but rather *too much fluency*, as 'easiness' means that we tend not to use all of our mental capacities. This contradicts some existing DMI research that uses HCI approaches to evaluate an instrument's usability [204, 212]. The question, of this study, therefore, is whether disfluency can make DMI players perform at a heightened level, and, most relevant to this research, whether this would translate into effects perceivable by the audience.

8.1.2 Motivation: Keith Jarrett and *The Köln Concert*

Over the course of this work I have engaged continuously with the function and nature of error. A concept closely related to error is the idea of *risk*.

Risk, like error, tends to lack rigorous definition, but the dictionary can be a place to start: '(Exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance; a chance or situation involving such a possibility' [2]. Risk, then, is a state in which something unwanted is more likely to happen.

Kaplan and Garrick make a distinction between risk and uncertainty that is relevant here. They state that risk 'involves both uncertainty and some kind of loss or damage that might be received' [113]. In live musical performance, this risk is often error. In looking for a way to understand how these related concepts of risk and error relate to music performance I came across the story of Keith Jarrett's performance of *The Köln Concert* [104].

Ian Carr retells this story in detail in his book on Jarrett's life [44, p. 71-73]. Jarrett, a master of jazz improvisation, was to play the first-ever jazz concert at the Opera House in Köln in January 1975. The organiser, 17-year-old Vera Brandes (Germany's youngest concert promoter), had ordered a piano but the wrong one had been delivered. The one that arrived was too small, and in disrepair: The notes stuck, the pedals didn't work, the felt had worn away in places, and it wasn't designed to produce the kind of volume a space like the opera house needed. Jarrett initially refused to play and was later cajoled into performing by Brandes, as a 1400-person sell-out crowd was due to arrive in a matter of hours and procuring another piano wouldn't be possible.

Jarrett was already an astounding improvisational performer, but in this instance his compositional choices were responses to the constraints of the instrument. The piano's high notes were tinny and the low notes not resonant, so he avoided the outer registers and stuck to the middle of the keyboard. He used rolling left hand riffs to compensate for the lack of bass notes, and to create a kind of sustain instead of relying on malfunctioning pedals. He stood at the piano in order to force enough volume out of it to reach the back rows. The result was a smash hit performance the audience loved, that has endured as a hugely popular recording for nearly 40 years.

More recently Tim Harford has used this story a hugely popular TED talk to argue for the value of messiness and frustration [91], suggesting that contending with unexpected problems can inspire us to new heights of creativity. But, any person simply sitting down at a piano will not produce such magical results. It seems that *The Köln Concert* is a merging of two serendipitous circumstances that combined to create an optimal creative opportunity: Jarrett's masterful facility with piano improvisation, and a substandard piano that had limitations that were unusual, and challenged him in particularly fruitful ways. In other words, desirable difficulty.

We don't go to musical concerts to watch people mess up or fail, and people didn't come to hear Jarrett play because they heard the piano was terrible and wanted to see what happened. What this concert's enduring popularity might signal is that when a performer is engaging with their own skill, their musicianship, understanding, and personal style is appreciable.

In DMI research, Fyans and Gurevich, in a study examining how audiences understand skill in a DMI performance context, found that spectators reported that the performer 'failing to engage with [the DMI] in an embodied way was indicative of a lack of skill' [72] — in other words, if the audience can't see the performer doing something, they don't perceive the performance to be skilful. This finding is reinforced by research in cognitive psychology, most notably Kruger et al.'s finding of the inverse: That audiences report higher ratings of quality, value and liking on work that they perceive to require more effort (also known as 'the effort heuristic' [122]). If effort is a marker for quality, it makes sense that, as Fyans and Gurevich found, that a perceived lack of effort would suggest a lack of skill.

In this way, disfluency, and its associated risk state, has the potential to be useful design tool for exposing skill.

8.1.3 Constraint and appropriation as distinct from disfluency and risk

There is existing knowledge on constraint and appropriation, and it is useful to disambiguate these from the notions of disfluency and risk.

The usefulness of constraint in DMI design to encourage personal playing style has been highlighted by Gurevich at al. [88] and reflected upon by Magnusson [131].

Zappi and McPherson [214] have demonstrated that constraint as a useful way to take advantage of appropriating behaviours. Appropriation, proposed earlier by Dix [61], is the notion that the people who use things will develop ways to use them that the designer neither intended or anticipated.

Zappi and McPherson tested the effect of constraint on performer action by presenting participating musicians with an extremely constrained interface, and observing the ways performers used appropriating behaviours to contend with it [214]. They depart from Dix's notions of appropriation [61], because appropriation is seen repeatedly throughout music history: The authors note prominent examples, such as distortion being an engineering headache until it was used by electric guitarists.

The authors found, when performers were presented with instruments that had extremely limited functionality, that players developed new playing behaviours in response. Further, the performers found this process to be creatively interesting and satisfying, thereby making the compelling case that the challenge of a limitation is useful.

Constraint on the surface appears to be the same as disfluency: Like the disfluency, constraint is a limitation around which performer develops behaviours to contend with it. However, the difference between constraint and disfluency is their stability over time; disfluency will diminish as behaviours to contend with it become second nature, therefore becoming less risky.

In this way, if we accept a definition of improvisation as 'the realisation of action as it unfolds' [115, p. 1], then dealing with an unstable disfluent quantity can be considered a type of improvisation. This separates disfluency from a constraint around which a performer can develop a style that can be practised. Performers can perform this and other behaviours in ways that are risky in order to challenge themselves and enter into a risk state, but constraint is by no means an automatic indicator of the presence of risk.

The risk that arises from disfluency, then, is not a design quality or physical element, but a continuous state. In a risk state, a performer will be ascertaining the limitation and may be performing this appropriation in real time.

8.1.4 Lessons from jazz improvisation

Following on from the story of *The Köln Concert*, ideas about jazz improvisation provide clues about how risk may operate in a DMI context. Most important of these is the notion of **previous improvised lines as enemy**. Jazz critic Francis Davis sums up the challenge of this aspect in this way:

> *The danger with that practice is that the music can become a progression of cliches. I mean, you go from doing one style you know how to do well to another style you know how to do.* [52, p.8]

Keith Jarrett echoes this in discussing his own practice as a solo concert improviser:

> *[M]y challenge in solo concerts was ... not to come up with good music I had come up with before.* [64]

Jarrett has clearly dedicated a lifetime of musical development to a temporal process of composing and performing, and has a deep understanding of what it means to be in the moment. Simply presenting any musician with a broken piano will not create good music, because it is not simply the piano that made the *The Köln Concert* an exceptional musical event. Instead, it is the combination of the instrument's limitations and Jarrett's phenomenal skill that converged to create the opportunity for optimal creativity.

Disfluency research supports this. Bjork and Bjork, leaders in the field, lend nuance to the findings around disfluency by specifying that it is not the simple presence of disfluency that brings about this state, but rather it is when disfluency that manifests as 'desirable difficulties' [28] brings one into this cognitively heightened space.

8.1.5 Summary: Disfluency as a tool for risk?

Disfluency is a condition of processing difficulty that has been found to engage higher cognitive processes. In light of the story of Kieth Jarrett's *The Köln Concert*(where a skilled performer contending with a broken piano produced a masterpiece), as well as existing investigation by Fyans et al. of spectator perception of skill (where they suggest that a lack of embodied engagement equates to a lack of perceived skill), disfluency is a potentially fruitful area for investigation.

To investigate if disfluency and risk produce useful effects for the audience, this study has the following central question:

> What are the relative effects of disfluency in a DMI on audience perception as it relates to audience enjoyment?

Additionally, this third study also lends insight into three of the questions that guide this thesis:

RQ2: What is the impact of visible risk on audience perception of DMI performance?

RQ3: Can the physical design of a DMI affect the performative outcomes?

RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

8.2 Study design: Instruments, players, context, and method

This study was performed using a DMI designed for the purposes of this investigation, called Keppi. Detailed technical and design information is in Chapter 6, but this section will detail how disfluency was incorporated into the design.

Then, this section also details the other design aspects of this study: The performers, the context of the study and the study method, as well as procedures of data collection.

8.2.1 The instrument

Keppi was the instrument designed for this study. An in-depth description of the design process of Keppi is found in Chapter 6.

Unlike Study 2, which studied the difference in audience perception of performances on 2 versions of the same instrument that had differing physical characteristics, each version of Keppi was identical but the instruments differed based on *behaviour*.

The design process of Keppi is described in depth in Chapter 6. The core of the disfluent behaviour was that the instrument needed to be continuously moved in physical space, and would turn itself off via a countdown if not moved enough. This expiring time was visible to the audience through 5 rows of lights that ran up the instrument.

This state was managed according to the quantity of motion generated by the performer's handling of Keppi, sensed by an accelerometer inside (see Chapter 6). If the quantity of motion lapsed, the lights would begin to tick down. When the motion had ticked down to the last row, the sound was distorted, and when the lights all turned off the instrument would stop producing sound. If entered into this 'off' state, Keppi needed to be shaken in order to charge it up again.

6 Keppis were were produced, 2 of each of the 3 test conditions:

Category 1: Control version Didn't contain an accelerometer, lights on continuously

Category 2: Mild disfluency Lights ticked down slowly (1600ms between ticks)

Category 3: Heightened disfluency Lights ticked down quickly (800ms between ticks)

The instruments were randomly assigned to the performers.

8.2.2 The players

For this study I looked for percussionists with experience of 5 years or more. This was to ensure that the performers could be assumed to have demonstrable skill in percussion, and that any audience ratings could not be dismissed as being because the performer simply wasn't very good at percussion.

I was fortunate enough to recruit 6 experienced percussionists: 4 Guildhall alumni, and 2 other skilled percussionists from personal networks. Each player was given one version of Keppi three weeks before the concert, and asked to prepare a composition that was three to five minutes in length. The percussionists were asked not to discuss the instrument with their peers, and were not told the details of the study or how the instruments differed, or what Keppi state they received. They were given the instrument, shown how it worked, and invited to compose and perform on it in whatever way they felt resonated with their playing style.

8.2.3 The study context and method

I used the same concert venue for this final study as for the previous two, which is the Film and Drama Studio at Queen Mary University of London. This is a 65-seat performance space with raked seating and a level stage. The study took place on May 24, 2017.

As with previous studies, this study was the middle act of an evening concert, which was on the theme of Music from the Augmented Instruments Lab[1]. The concert was advertised by email lists and social media. There were three acts, each act featured work from the research group: First, Jack Armitage live coding; second, the six Keppi performances; and finally, a performance on Chimney, an interactive performance interface designed by Fabio Morreale, by Raul Masu featuring Naomi McLean on cello.

After the first act I introduced the study to the audience and invited them to participate. I explained how data was gathered, and survey books were distributed (the questionnaires used are available in Appendix M) and then screened the onboarding video for Metrix. In total, 39 audience members participated. All audience members watched all performances, and there was no split into audience groups, as the axis of comparison does not lie in the audience but rather in the difference in the performers and the behaviours of their instruments.

The order of performers (detailed in Table 8.1) was randomly assigned, and by chance each half of the performances had one of each DMI condition.

The performers stayed outside until their time to play, to avoid them being influenced by the playing style of the others. When it was their turn to play they were brought into the performance space and introduced to the audience with their name.

Questionnaire design

The questionnaires for this survey were similar in form to those of the previous studies (they are available in Appendix M).

[1]Original news post: `http://instrumentslab.org/news/events/2017/05/21/music-from-the-lab.html`

Table 8.1: Order of performances and instrument states for each

Performance	DMI behaviour category
Performer A	Category 3 (heightened disfluency)
Performer B	Category 1 (control)
Performer C	Category 2 (mild disfluency)
Performer D	Category 1 (control)
Performer E	Category 3 (heightened disfluency)
Performer F	Category 2 (mild disfluency)

Though the post-concert survey was largely unchanged from those used in Studies 1 and 2, two changes were made to the post-performance survey. Two changes were made from previous studies. First, the qualitative ratings of Interest and Understanding were removed, because I was primarily interested in Enjoyment. Additionally, I removed the qualitative question 'How would you describe this performance to a friend?' in the interest of brevity.

8.2.4 Notes on technical problems

There were some unanticipated technical problems during this study. First, I had planned to amplify the sound of Keppi as its small in-built speakers are not loud enough for the space. I had intended that this would take place via a radio microphone placed inside the instrument that would broadcast to the space's PA system. However, the radio mics ended up being extremely noisy, and this solution was simply untenable. We compensated for this by placing condenser microphones on stage that would pick up the sound and amplify it via the PA system. The performers were allowed to decide before the concert how they would set themselves up on stage, and the microphones did not limit movement or gestures.

Secondly, there was a persistent bug in the capacitive sensor system. So far it's unknown whether this is a function of the unusually large electrodes used in Keppi, or if it is a problem that arises because of the difference in skin conductance between individuals, or if it was an issue of the sensor scan rate (which runs over I2C), or if it was a function of the placement of the piezo sensors or due to manufacturing differences in the discs themselves. The result was that, at times, hitting the instrument didn't produce sound. During development this issue seemed subtle and like it would be rare, but it was noticeable by the audience (this is discussed in Section 8.3).

Finally, there was one instance of malfunction during Performance E. I had suggested to the performers that to keep the internal hardware stable they might stuff the instrument with something to pad the inside. All did this, but one performer's instrument, that had been stuffed with socks, turned itself off in the first minute of her performance. This was the fifth of six performances, and may be attributed to the instrument being on for an extended period of time through sound check and then through the first act of the concert, and over half the second act. The performance was stopped and the performer started it again, using the other instrument that had the

Table 8.2: Column statistics for data set, including mean Enjoyment rating for each performance.

	Perf. A	Perf. B	Perf. C	Perf. D	Perf. E	Perf. F
Mean rating	3.576	3.788	3.152	4.455	4.03	4.333
Std Dev	0.8303	0.8572	0.9395	0.6657	0.9515	0.6922
Std Err of Mean	0.1445	0.1492	0.1635	0.1159	0.1656	0.1205
Median	4	4	3	5	4	4

same risk condition as the one she was using.

Though these problems existed, they were not contained to any one performance, and therefore no performer was unfairly advantaged or disadvantaged. In the case of the instrument malfunction during Performance E, though it is impossible to know precisely how this impacted audience perception, the performance was still one of the most highly rated so we can conclude that it was not overly detrimental.

8.3 Results: Post-hoc

Out of 39 audience members, 31 provided complete data sets. Therefore, in the post-hoc and real-time results sections, N=31.

8.3.1 Quantitative

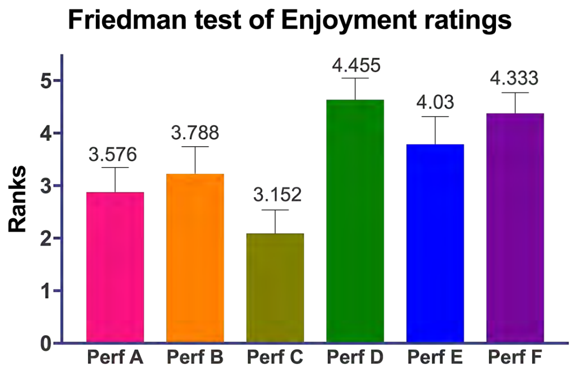
The quantitative data of this study is on two aspects: The ratings of Enjoyment for each performance, and the rank ordering of the performances at the end of the concert.

Enjoyment ranking

A Friedman test was used to analyse the ratings of Enjoyment for each performance, which were on a scale of 1 (not at all) to 5 (very much). The Friedman test was chosen for the following reasons:

- The data was non-Gaussian (according to a D'Agostino & Pearson normality test, α=.05);
- There was only one factor (the Enjoyment rating);
- The subjects were independent (the audience members did not discuss their scores with each other);

Figure 8.1: Friedman test of Enjoyment ratings. Data labels indicate mean rating. Error bars indicate 95% CI.



- There were 3 or more groups (6 performers in total);
- The ratings across the 6 performances were paired, because the same audience members judged all 6 performances.

A Friedman test was run to determine if there were differences in ratings of Enjoyment across the six performances. Pairwise comparisons were performed with a Dunn's correction for multiple comparisons. Ratings of enjoyment were statistically significantly different at the different time points during the exercise intervention, $\chi^2(5) = 52.31$, p<.0001.

Post hoc analysis revealed statistically significant differences in ratings of Enjoyment between the following performances (statistical significance was accepted at the p<.05 level):

- Perf. D rated significantly higher than Perf. A (p=.002), Perf. B (p=.033) and Perf. C (p<.0001)
- Perf. E rated significantly higher than Perf. C (p=.0034)
- Perf. F rated significantly higher than Perf. A (p=.0169) and Perf. C (p<.0001)

To examine the ratings of Enjoyment according to instrument type, Wilcoxon signed-rank tests were performed on three pairings (of instrument condition) to determine if there were statistically significant differences between the ratings based on instrument type. Data stated are media values, followed by mean values.

A Wilcoxon signed-rank test performed on the Enjoyment ratings for the Control condition (Med=4, mean=4.121) and Condition 1 (Med=4, mean=3.742) found the differences between the paired ratings to be statistically significantly different (p=0.0057).

A Wilcoxon signed-rank test performed on the Enjoyment ratings for the Control condition (Med=4, mean=4.121) and Condition 2 (Med=4, mean=3.803) also found the differences between the paired ratings to be statistically significantly different (p=0.0211).

No statistically significant differences were found between the paired Enjoyment rankings of the Cat. 2 and Cat. 3 instruments.

Rank ordering

In the post-concert survey, the audience was asked to order the performances in order of preference, from 1 (favourite) to 6 (least favourite). a Friedman test was performed on these results to test for significance.

Pairwise comparisons were performed with Dunn's correction for multiple comparisons. Rank ordering was statistically significantly different for various performances during the exercise, $\chi^2(5)$=31.46, p<.0001.

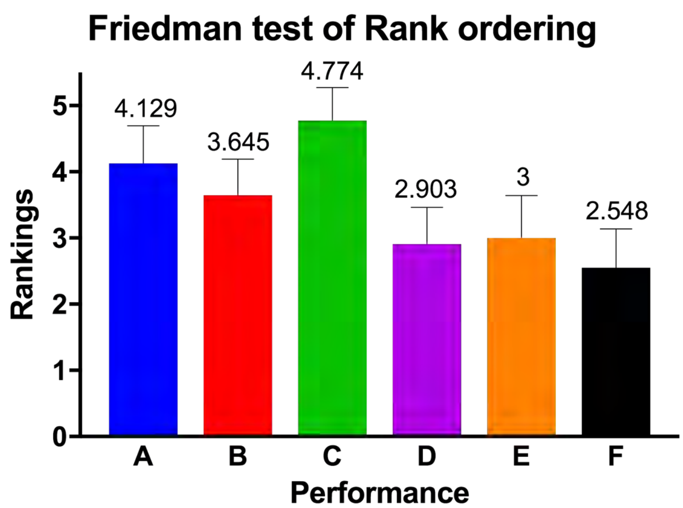


Figure 8.2: Illustration of Friedman test of rank ordering. Data labels indicate mean values. Error bars indicate 95% CI. NOTE: Since the rank was from 1 (favourite) to 6 (least favourite), shorter bars indicate stronger preference.

Table 8.3: Mean rank value, standard deviation, standard error of mean, and median value of Perf. A-F.

	Perf. A	Perf. B	Perf. C	Perf. D	Perf. E	Perf. F
Mean	4.129	3.645	4.774	2.903	3.00	2.548
Std dev	1.544	1.496	1.359	1.513	1.751	1.609
Std err of mean	0.2772	0.2686	0.2441	0.2718	0.3145	0.289
Median	5	4	5	3	2	2

Post-hoc analysis revealed statistically significant differences in rank ordering between the following performances[2] (statistical significance was accepted at the $p<.05$ level):

- Perf. A was preferred significantly less than performance E ($p=.462$);
- Perf. C was preferred significantly less than Perf. D ($p=.0004$), Perf. E ($p=.0010$), and Perf. F ($p<.0001$)

8.3.2 Qualitative

I performed a thematic analysis on the qualitative results of the responses to the question of 'What did you like about the performance?' and 'What did you dislike about the performance?' in the six post-performance surveys. The process used is described in Section 5.4.

After familiarising myself with the data, I then coded it, resulting in 54 codes. These were then grouped into themes. (The codes, the way they were assembled into themes, and the code/theme distributions are available in Appendix N.)

Themes

The four themes present in the other two studies — Ways of Playing, Instrument, and Sound — emerged in the qualitative data corpus for this study. In addition, there was a new theme: Skill. Mentions of skill were present for the 'like' responses of Performances A, D, E and F. Mentions of skill were present for the 'dislike' responses of Performance C.

[2]An important note about rank ordering: Audience members ranked the performances from 1 (favourite) to 6 (least favourite). This means that **lower** values indicate performances that were less preferred. This can result in counter-intuitive statistical results, i.e. the less-preferred performances seem to have 'higher' mean ranks; however, preference is indicated by lower means ranks, as these ratings are closer to 1 (favourite).

Differences between test conditions

The qualitative data themes were compared on the axis of test conditions. The performances with instruments with the same behaviour were Performances B and D (Cat. 1), Performances C and F (Cat. 2), and Performances A and E (Cat. 3).

Though all three sets of performances had 'enjoyment' events on the theme of Player Action, the control instruments did not have events that involved physicality, whereas the other instruments did.

On the predominant themes in the coded 'error' events, differences were also observed. Though all instrument types had 'error' events related to rhythm inconsistencies and trivial technical errors, the performances using instruments with control conditions had 'error' events predominantly related to judgement of performer action, such as hesitation and intention. (Cat. 1 also had 'error' events on this theme around intention, but this was at the beginning of the performance when the player was demonstrating the instrument's capabilities, and not during her performance.)

8.4 Results: Real-time

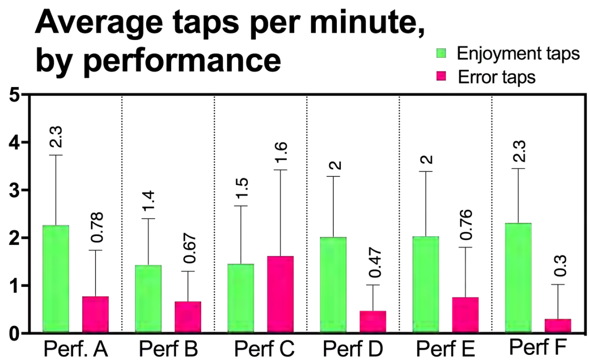
Real-time data was collected using Metrix. As for Studies 1 and 2, the Enjoyment tap rates are expressed in time bins of 5s and the 'error' tap rates are expressed in time bins of 1s. For each participant, any more than 1 tap per time bin were discarded. For correlations and averages, however, all taps were considered.

8.4.1 Observations from the time series data

The real-time data was visualised to give an overall sense of audience reaction during the performances. These histograms are available in Appendix K.

The real-time data from this study shared several features with the real-time data from Studies 1 and 2. Firstly, there were again more 'enjoyment' indications than 'error' indications over all performances. Secondly, the peaks in the data are different; 'enjoyment' peaks tend to be cumulative, building to a peak and then tapering off again, whereas 'error' peaks have a sudden appearance and drop-off.

Figure 8.3: Visualisation of average number of button taps per minute, by performance; values shown. Error bars indicate SD.



As well as these patterns in visualisation, the tapping behaviour of individuals, much like in Study 1 and 2, remained fairly consistent. Enthusiastic button tappers maintained this usage pattern, and conversely those who were much more conservative with their feedback were also consistent in this respect.

8.4.2 Normalising the 'enjoyment' and 'error' tap totals

In this study, the duration of the performances in this study varied considerably, from 1:31s to 3:59s. A straightforward comparison of the number of button taps per performance would therefore be misleading, as some performances left much more time for the audience to make judgements. As such, I have normalised these values as an average number of taps per minute of performance (see Figure 8.3).

Participant data was included in this set if the participant had used one of the buttons at least once during the performance, to ensure that any zero values came from someone who was active and not someone who had not simply ceased participating.

8.4.3 Differences between rates of 'enjoyment' and 'error' taps

To examine the differences between the number of 'enjoyment' and 'error' taps, the normalised tap-per-minute data set described in the section above was used.

A Kruskal-Wallis test was run on these 'enjoyment' and 'error' tap rates to assess whether there were any significant differences in per-minute tap rates between the 6 performances. Pairwise comparisons were performed using Dunn's procedure with Dunn's correction for multiple comparisons. Adjusted p-values are presented, and the results are illustrated in Figure 8.4.

This post-hoc analysis found no statistically significant differences between the rates of 'enjoyment' taps between any of the six performances.

For 'error' taps, one difference was found that was statistically significant. Performance C's rate of 'error' taps was statistically significantly higher than that of Performance D ($\chi^2(5)=41.15$, p=.0187) and Performance F ($\chi^2(5)=58.98$, p=.0001).

This analysis was also run on tap rates divided by instrument condition. A Kruskal-Wallis test was run on these 'enjoyment' and 'error' tap rates to assess whether there were any significant differences in the per-minute tap rates between the three types of instrument. No statistically significant differences were found (see Figure 8.5).

8.4.4 Correlation with post-hoc Enjoyment ratings

Correlations of tap rates and Enjoyment ratings were done across all 6 performances. This data set included participants according to two criteria:

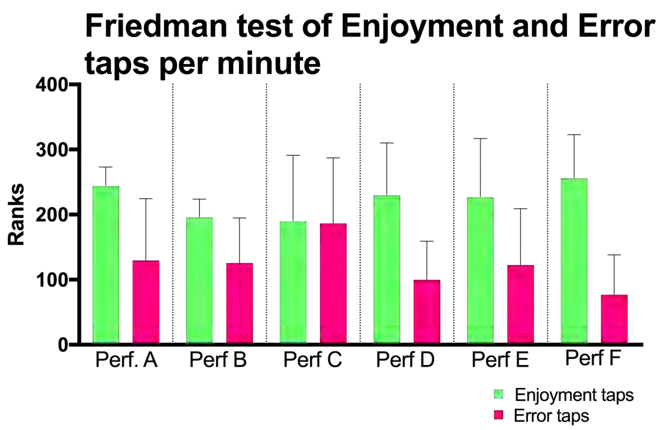
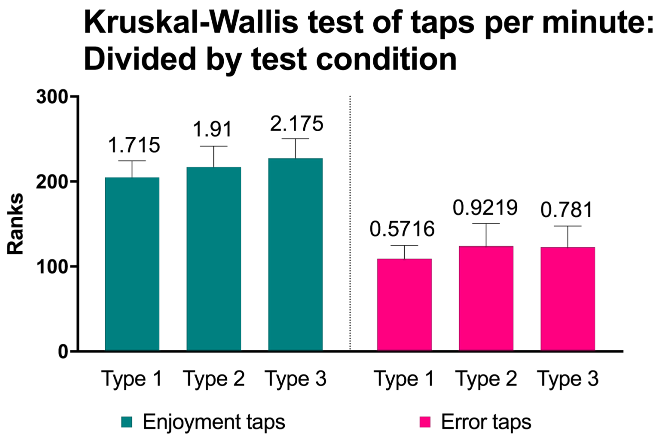


Figure 8.4: Plot of Kruskal-Wallis test on taps per minute per performance. Tap values were normalised to per-minute values. Error bars indicate 95% CI.

Figure 8.5: Plot of Kruskal-Wallis test on taps per minute divided by the three test conditions. Tap values were normalised to per-minute values. Error bars indicate 95% CI. Data labels indicate mean taps per minute.



First, that they had provided a post-hoc Enjoyment rating, and second, that they had used one of the buttons at least once during the performance.

A Spearman's test was used to assess the relationship between the average number of button taps per second the participant made, and the Enjoyment rating they gave. Out of 12 comparisons made, there were two correlations of statistical significance found.

For Performance C, there was a statistically significant positive correlation between the number of 'enjoyment' button taps and the resulting Enjoyment rating (r_s=.45, p=.04). For the same performance, there was a statistically significant negative correlation between the number of 'error' button taps and the resulting Enjoyment rating (r_s=-.60, p=.003).

Using the same criteria, the tap rates and Enjoyment ratings were grouped by test condition, and these tap rates were correlated with the ratings of Enjoyment for each instrument type. No positive or negative correlations of any statistical significance were found for the test condition groupings.

8.4.5 Identifying 'events' in the data

The following processes were followed to find events in the time series data (for reference, the grouped events are highlighted in the histograms in Appendix O):

Enjoyment events: The threshold of audience agreement in a given 5s time bin for the start of an 'enjoyment' event was 6. The event starts from this bin, continues to the peak, and ends on the next bin until audience agreement is 5 less than the peak, or is less than 6. It is acceptable for events to start and end on the same bin.

Error events: The threshold for error events in a given 1s time bin was 3. An event starts 1 bin before this (if applicable), and continues until there have been no error events for 2s.

Histograms of the real-time data with the marked event ranges are included in Appendix O.

8.4.6 Analysis of video data

Links to the video documentation can be found in Appendix P.

I coded the video by reviewing the video documentation at times of 'enjoyment' and 'error' events, determined using the procedure outlined above. This was done using the process described in Section 5.5.7. A summary of the video coding, as well as the theme groupings, are also available in Appendix P.

8.4.7 Themes of 'enjoyment' events

The codes generated from the video documentation at the times of 'enjoyment' events clustered around the following themes:

1. **Novelty:** Change, new elements.

2. **Pattern:** Rhythm, repetition.

3. **Player action:** Intention, technique/skill, physical actions.

4. **Time-based features:** Establishment of themes, moments of flow, build-up.

The highest-rated performances (Performance D, E and F) all included moments related to player action, and specifically to skill — such as fast, complex rhythms.

The lowest-rated performance, Performance C, had enjoyment events clustered around time-based features such as the establishment of rhythm.

However, this performance only had 3 ‘enjoyment’ events that met the threshold. Performances A and B had enjoyment events that mostly featured change, such as a new element or method of interaction.

8.4.8 Themes of ‘error’ events

The codes from the ‘error’ events were clustered around the following themes:

1. **Judgement of performer action:** Moments of hesitation, unsure, unclear intention
2. **Interrupted expectations:** Breaks in rhythm, changes, unexpected elements, sound distortion
3. **Trivial technical errors:** The performer adjusting the instrument, gestures that do not make sound, moments of struggle with the instrument.

The performances all had between 9 and 14 ‘error’ events, with the exception of Performance F, which only had 3. ‘Error’ events tended to be predominantly related to themes of interruption, such as moments where the rhythm was inconsistent. The theme of performer action was also salient, and moments where performers were unsure, or hesitated, tended to be marked as errors. Technical errors such as the instrument not sounding or the performer not being confident and fluent in their handling of the instrument also tended to be considered errors.

8.5 Discussion

This section discusses the findings of this study as they relate to the research questions. I will first discuss the findings related to the central study question, and then discuss how the results of this study relate to the larger questions of this thesis.

8.5.1 Findings related to the central question

This study aimed to determine if a risk state, incorporated through disfluent instrument behaviour, would have an effect on audience perception. In Keith Jarrett’s recording of *The Köln Concert*, where the limitations of a broken piano presented a *useful limitation* that, when he used his remarkable skill to contend with it, gave rise to a genius improvisational performance.

Departing from this idea of disfluency being negotiated using skill, this study presented seasoned percussionists with the same instrument with one of three levels of disfluency and risk: None (control), mild disfluency, and heightened disfluency. The central question of this third study was as follows:

> What are the relative effects of disfluency in a DMI on audience perception as it relates to audience enjoyment?

The quantitative data showed no clear indication that disfluency had a statistically significant impact on audience ratings of Enjoyment. Two performances with similar mean ratings, Performance D and Performance F, were Cat. 1 (control) and Cat. 2 (mildly disfluent) instruments. These instruments were also the most preferred in the rank ordering.

This was an unexpected result and there are two possible reasons for this. First, the study design may not have isolated the effects of disfluency. Though I did not want to simply make the instrument hard to play, allowing the performers time to use their instrument before the concert could mean they gained a degree of familiarity with it, cancelling out any effects of disfluency. Perhaps if they had each been given a control instrument and this was swapped for a disfluent instrument shortly before the performance this would not have been the case.

The second reason is that it may not be the degree of visible risk that is important to audiences. The qualitative data provided deeper insight here. The four themes seen in the qualitative data of Study 1 and 2, Sound, Way of playing, Instrument, and Experience were present, but there was also a new theme: Skill. For Performances A, D, E, and F skill was mentioned in a positive sense. For Performance C — the performance that, in the qualitative data, was significantly lower rated than others and less preferred — skill-related *dislikes* were present.

Performer C had a Cat. 2 instrument (mildly disfluent). In the video documentation the audience indicated many 'error' events, and in review of the video he is noticeably less fluent with it, holding it in one hand and playing with the other, without any of the creative physical playing methods employed by the other performers. It may be that, in this case, disfluency had a detrimental effect. Rather than being a useful limitation — one that matched his ability, and that he could leverage his skill to contend with — it may simply have been a limitation, and this was perceived by the audience as a deficit of skill.

For the other 5 performers, their skill and the type of disfluent behaviour were well matched, and they enjoyed and rose to the challenge. The fact that one performer did not cope well with it does perhaps suggest that the mere presence of disfluency is not a creative panacea. It has to be the right kind of disfluency, matched to the right level of skill. This is similar to Cskiszentmihaly's work on *flow*, his central tenet being that, in order to reach a 'flow state', skill and challenge must be in balance [48].

It is also important here to revisit the time-based nature of disfluency. I positioned disfluency as distinct from appropriation earlier in this chapter by noting that disfluency is a time-based quality, a limitation that the performer has to use their skill to manage. By this definition, it is probable that disfluency will also diminish over time: The disfluent feature and all its effects will become known, strategies and tropes will be developed, and behaviour that was once disfluent will become a stable design feature.

It may be that the results of this study do not show any significant difference in the qualitative ratings between versions of Keppi because I gave the performers time to practice. This might, in retrospect, have been mitigated in a number of ways: I could have given all performers a control instrument to practice with, and switched this for a disfluent instrument before performing. This may make skill even more noticeable, or it may prove distracting to the audience. It is impossible to speculate, but this is certainly an intriguing question that warrants further study.

8.5.2 Findings related to the larger research questions

Visibility of risk

> *RQ2: What is the impact of visible risk on audience perception of DMI performance?*

The lights on the outside of the instrument were intended to make visible the risk state of the instrument, and were active in Performance A, C, E and F. There was not evidence of the audience being specifically cognisant of the risk state of the instruments, and there was no clear preference for a ‘risky’ instrument over a ‘less-risky’ one (in fact, the two highest-rated performances were Performances D and F, and D was a control instrument).

Though risk being perceivable and understandable to the audience does not seem to be necessary, in this study audience was remarkably sensitive to the presence of skill. They recognised and commented upon the skill in five of the performers — a qualitative theme not present in the data from Studies 1 and 2 — and they also picked up on the remaining performer’s relative lack of experience. This supports Fyans and Gurevich’s definition of skill [72] as an embodied phenomenon that is a combination of control and effort.

Perhaps, then, what produces a space of musical possibility where some outcomes are more interesting than others is not visible *risk*, but visible *skill*. The implication for DMI designers, then, is that instruments that offer the opportunity for displaying effort and control would be useful for audiences.

The influence of the design on performative outcomes

> *RQ3: Can the physical design of a DMI affect the performative outcomes*

The results of this study indicate that while the audience did not appear to respond not to visible risk, they responded to visible skill. The question of the role of the DMI in these outcomes, then, becomes one of how the DMI’s physicality can help make skill perceivable.

Keppi was designed to encourage interactive pluralism. Although it had to be struck to make a sound and had a limited sound palette, that striking could take place in a wide variety of ways. Keppi was also intentionally round, to not only make it conducive to being gripped but also so the most obvious way of playing it would not be to place it on a table or a stand. The disfluent behaviour further encouraged performers to move the instrument, which required them to engage their motor skills and physical coordination, and to develop different playing strategies.

The design of Keppi was approached with interactive pluralism in mind (discussed in Chapter 6). A notable feature of these six performances is the diversity of playing methods that the performers displayed. Performers struck it with hands, fingers, feet, and bounced it on a knee. Performance D, a control performance, devised a way of balancing it on two chairs de-

Table 8.4: Description of playing approach by each of the six performances. A variety of playing styles and methods were used.

Performance	Description of playing
A	Played standing, striking with her hands, and incorporated a complex counter-rhythm by stamping her feet.
B	Played seated, and demonstrated different methods of playing with hands, fingers, and feet.
C	Played standing, holding with one hand and hitting with the other. Incorporated a shaking action.
D	Balanced the Keppi between the backs of two chairs and played it with her fingers like a piano; displayed dynamics and subtlety
E	To keep the instrument moving she bounced it up and down on her knee and also added rhythms with her hands, which produced a rhythmic drone; style was aggressive.
F	Incorporated the most physical movement with the instrument, played with a combination of slaps and catches. Made use of the tick-down behaviour as a structural element.

spite it being round, enabling a performance of rapid finger-tapping — a strategy that would not have been possible if their instrument had disfluent behaviour. Performance A, by an expert in complex rhythms of Ghanaian percussion, created an intricate counter-rhythm with her feet. Performance E featured a drone created by the performer bouncing the instrument on her knee to keep it moving. Performance F incorporated a swinging motion. Performances A, E and F — all with disfluent instruments — incorporated the disfluent behaviour, or the tick-down of the lights, into the performance.

This diversity of approach suggests that interactive pluralism was effective, and produced a wide range of playing approaches (see Table 8.4). Additionally, in an informal survey after the study, performers were asked what they liked about the instrument and all players who had a version of Keppi with disfluent behaviour cited this as a positive aspect:

- ‘I liked that I had to be active to make it work.’
- ‘The necessary movement leads to [movement] in rhythm.’
- ‘I liked the limitation set on it where it had to be constantly moved. I've always found that limitations result in increased creativity on the performer's part.’
- ‘I did like the aspect of it having heft, and needing to be explored. I enjoyed that it took some effort to play and come to grips with.’

Along with interactive pluralism, the disfluent behaviour also played a role in this diversity. Contrasting the control performances with the performances on disfluent instruments, there were a number of strategies that may not have been employed were it not for the requirement to keep the instrument moving (such as the knee-bouncing in Performance E, or the swinging of the instrument in Performance F).

Combining the post-hoc and real-time data

> *RQ4: How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?*

This study is the third use of the combined methodology that is a key contribution of this thesis.

While the quantitative audience ratings did not produce any clear results, the qualitative and real-time data uncovered that visible skill had an effect on audience perception, and being able to query the video documentation at the time of 'enjoyment' and 'error' events was once more an important factor. This is a further indication that one data set does not tell the whole story, and that being able to integrate post-hoc and real-time data can lead to more nuanced findings.

8.6 Summary

This chapter described the third study of this thesis, which queried the relative effects of disfluency in a DMI on audience perception. Keppi, a percussion DMI, was designed for this investigative purpose (described in Chapter 6). 6 Keppis were produced, 2 of each of the 3 states: No disfluency, mild disfluency, and heightened disfluency. These were given to 6 seasoned percussionists, who composed a short piece that they performed for the study audience in the context of an evening concert.

This study did not find clear effects of disfluency on audience perception. However, in the analysis of the qualitative data and video analysis, a new theme emerged: Skill. Instead of creating a risk state perceivable by the audience, disfluency instead seems to create the opportunity for audiences to perceive and appreciate performer skill. The definition of skill from Fyans and Gurevich [72] as an embodied quality that combines *control* and *effort* is relevant here, as it might be that a disfluent instrument behaviour requires the performer to exert control and effort contend with it, and this skill is what the audience is picking up on.

Further, though interactive pluralism, as discussed in Chapter 6, was a primary design value, the diversity of technique by the performers may be due also to the disfluent behaviour of the DMI. To isolate the influence of disfluency on the performer and in turn on the perception of the audience, however, requires further study.

This was also the third iteration of the combined post-hoc and real-time methodology. Where the quantitative ratings of Enjoyment did not show any clear preference, the real-time indications of ‘enjoyment’ and ‘error’ reinforced the qualitative findings, lending weight to the finding that the audience values visible skill.

In the following chapter, Chapter 9, I provide a review of the four central study questions of this thesis, and discuss in detail how the findings of this and the previous two studies provides answers and insight.

Chapter 9

Discussion

This thesis was guided by four research questions that centre around the audience perception of DMI performance. In sections 9.1 to 9.4, I discuss what the results of the three studies in this thesis contribute to the four overarching questions and the implications therein for DMI designers, as well as how these results connect to existing work presented in Chapter 2.

I conclude this chapter by revisiting the central contributions of this thesis, and reflections on lessons learned in carrying out research in a live DMI context.

9.1 Audience perception of error in DMI performance

> **Question 1:** With no external stylistic frame of reference, how does the audience perceive error in DMI performance, and how does this affect enjoyment?

DMI performance presents two confounding factors to audiences: It is played with unfamiliar instruments, and in an unfamiliar music style that is often abstract and improvisational.

Both these factors have long been identified as problematic for audiences, as discussed in Chapter 2. Fels et al. [68] identify that the audience not understanding the mapping of gesture input to gesture output (which they term *transparency*) is problematic, and they suggest that an instruments' workings being 'opaque' makes DMI performance 'more difficult [for audiences] to appreciate'. Meanwhile, Fyans et al. suggest that, due to the lack of stylistic conventions typical of DMI performance practice, it was perhaps not possible to make errors at all [75].

Study 1, detailed in Chapter 5, aimed to isolate and examine these effects on audience enjoyment of DMI performance, and revealed relevant new insights into how audiences perceive error, as well as the roles that the instrument and the musical style have to play. Each half of the audience was given a technical tutorial on one of the instruments played in the study, and despite this technical familiarity there was no indication in either the post-hoc or the real-time data sets that having this knowledge produced any significant effect on audience enjoyment. This is not to dismiss the importance of transparency as proposed by Fels et al. [68]; it is possible that transparency may positively affect audiences, but these results indicate that transparency cannot be achieved by simply showing an audience how the instrument works — an important implication for the DMI community, as instrument demos have become commonplace as a way to mitigate a lack of transparency, and these appear to have little effect. A deeper understanding of the nature of transparency and its effects is an intriguing avenue for further study.

9.1.1 Error and musical style

It is the comparison of audience perception between musical styles that offers the most insight into this first research question. The two performers each played two performances, one performance in an experimental style (abstract and improvisational), and the other in a conventional (vernacular) style. In Chapter 2 I discussed at length the historical origins of DMI's radically experimental musical style, particularly with respect to Goehr's notion of Werktreue [83]. True to the modernist beginnings, a prominent characteristic of DMI performance practice is its rejection of existing systems of musical convention, specifically the Western classical tradition, instead placing value on forward-thinking exploration. Lydia Goehr applies the label *Werktreue* to what she sees as a culturally dominant, historically rigid, and arguably hegemonic musical system [83], but she specifies that, despite all the radical experimentation, DMI practitioners do not entirely reject this system; they may reject its stylistic conventions, but they still retain the 'formal constraints' such as the role of the audience and the concert hall setting.

Study 1 replicated this: Care was taken to retain the formal constraints, and the contrast between musical styles allowed an axis of comparison, and resulted in two preliminary findings. Firstly, the real-time results of Study 1 do not confirm Fyans et al.'s suggestion that error may not exist where stylistic convention is absent. Instead, the audience did indeed notice errors, even in the experimental performances that did not conform to any stylistic *Werktreue*. Further, the number of errors the audience identified in the experimental performances was not significantly different than the number identified in the conventional ones. We can conclude, then, that a stylistic frame of reference is not necessary for audiences to perceive error.

The second finding sheds light on how these judgments might take place. The audience appears to be using different systems to judge error, depending on the musical style of the performance, and the video coding using the real-time data provided insight here. There were four themes of 'error' events: Abrupt changes (such as sudden loudness), performer action/inaction (such as hesitation, or facial expression), trivial technical error (such as bumping into something), and musical inconsistencies (such as imprecise rhythm or pitch). In the conventional performances 'error' events across all themes were present, but the dominant theme of error was musical inconsistency. In the experimental performances, 'error' events clustered around themes of abrupt change, performer action/inaction, and trivial technical error, and events on the theme of musical inconsistency were absent. These results indicate that, while the style did not prevent the audience perceiving errors, the audience did not use the same criteria to judge the performance they were watching.

The predominance of 'error' events related to musical inconsistencies in the conventional performances indicates that the audience was using expectations of pitch and rhythm, gained through a musical frame of reference, to identify 'error'. These expectations seem to have dominated perception, as most 'error' events were on this theme, and far fewer were related to themes such as performer action. This could be Goehr's Werktreue in action: Goehr states that Werktreue is a pervasive system of musical understanding rooted in Western classical ideals, and where the conventional performances conformed to Western musical expectations the audience seems to automatically applied this system. Additionally, the 'top-down' system from Narmour's I-R model, which describes audiences judging what they hear against their own expectations, may have application here. Narmour states that the audience's expectations evolve while they listen, and within this data set this appears to be true; the audience identified musical inconsistencies according to each performer's playing. This is an indicator that, while the audience is using a musical system as a frame of reference for conventional performances, it is not a system that is rigidly applied across both performers, as the audience seems to adjust it to judge the performance at hand. This raises interesting questions — though *Werktreue* may be broadly culturally present, how impactful is it to audiences? Further, how do we apply this knowledge to individual performances — does adjust our existing knowledge to suit whatever we are watching?

The results most relevant to RQ1, however, are how the real-time audience perception of the experimental performances differ. For the experimental performances, which did not have specific stylistic constraints, the audience still identified plenty of 'error' events, indicating that the existence of error does not rely upon a widely accepted stylistic frame of reference. This is not to say that audiences don't attempt to use prior musical knowledge to understand what they're hearing, but rather when this frame of reference is not useful they don't simply stop trying to understand. This aligns closely with Dervin's sense-making methodology (SSM) [54, 55], which proposes that, when met with a gap in their knowledge, people use some combination of their personal frame of reference and in-the-moment contextual data to make inferences. The audience in this study, as Dervin proposes, bridged their gap in understanding presented by an unfamiliar musical style with contextual information, gleaned from performer action/reaction, or events that seemed sudden and out of place, as well as their existing frame of reference that might exist as a function of what Goehr calls the 'formal constraints' [83, p. 264] of DMI performance. For example, they identified the performer bumping into something as an 'error', which suggests that they have expectations of how performers interact with their instruments on stage, and that bumping into something is accidental.

These findings raise interesting further questions about how we form musical expectations. Do we form expectations at a cultural level as a result of deep influence of one aesthetic system, as Goehr suggests? How do these expectations play out in the moment, such as in Narmour's top-down system? When these stylistic references aren't useful, does the audience abandon them completely? If so, how and when?

The most important implication for DMI researchers and performers from this study is this: That whether DMI music adheres to musical convention, or defies it entirely, the audience is always judging. If the musical style is entirely unfamiliar and does not conform to any system of reference or expectation, the audience will still try to understand, and will look to the context for meaning if they can't make sense of the music. As a result, contextual information such as gesture and facial expression become even more important methods of communication.

This highlights a further important implication for DMI performance practitioners. Without a frame of reference, audiences use their own experience and contextual data to make judgements. No performer has control over the past experiences of their audiences, but performers do have control over the context in which they perform, as well as their gestures at every degree of subtlety. These results conclude that, though the audience brings with them a frame of reference, the performer can greatly influence the audience through contextual cues, and these are particularly salient in experimental musical contexts.

9.1.2 How error affects enjoyment

Though the post-hoc quantitative data for Study 1 indicated that musical style had a significant effect on audience ratings of Enjoyment, these results can't tell us anything about the nature of error, or its relation to enjoyment. However, the real-time data set, made up of audience indications of 'enjoyment' and 'error', provides much insight about the relationship between these terms.

In the real-time data set for Study 1, it was notable that 'enjoyment' and 'error' events indicated by the audience were not diametrically opposed, and often occurred at the same time. Though there is often assumed to be a binary relationship between these two terms, with 'enjoyment' being positive and 'error' being negative, these results indicate that these two terms do not have the binary relationship and seem to be independent.

It bears mentioning that this lack of evidence for a relationship between 'error' and 'enjoyment' was not only found in Study 1, but it was found also in the results for Studies 2 and 3. This repetition takes this from an indication of a potentially interesting area of further study to the realm of evidence of lack of relationship between these two terms. It was particularly notable that Performance 4 of Study 1, which was rated significantly higher than all other performances for Enjoyment by the audience, also did not have a significantly lower number of 'error' events than any other performance. From this, we can deduce that errors were perceivable and noted by the audience, but the performance was still widely enjoyed. This is a particularly salient indication that error, instead of detracting from audience experience, may instead have a different function, and that musical experience is not a zero sum game; enjoyment is not rubbed out simply because an error occurs.

This presents two new areas of inquiry. Firstly, what is a better way to talk and think about error in this context? These results are a solid first step in better understanding this term, but in order to gain insight into error's role in musical performance this understanding needs to be made more robust, and we need better models for describing it that are not infused with value judgement and do not assume binary opposition. In Chapter 2 I discussed Claude Shannon's model of information theory, and this is perhaps a starting point; this model proposes though a transmission may have noise, this does not make the message worthless.

Applying this model to the function of error in musical performance is a possible next step, and one for which, regrettably, the timeline and scope of this thesis did not allow. It's clear from these results that even a many errors do not render an enjoyable performance unenjoyable, but is there a point at which error interferes with, or even eclipses enjoyment? Is there some kind of event horizon where errors are sufficient in scale and/or number to make a performance irrecoverable — a fiasco point? Does there exist a diversity of error types that affect audience perception in different ways? How would a model apply when there is no solid stylistic end goal, such as experimental DMI performance?

The second further area of study is understanding how error intersects with personal style in a DMI performance context. Gurevich [88] suggests that diversions may be the seat of personal style, but in experimental music there is usually no score or system of understanding from which to diverge. What does this mean for personal style? The themes of 'error' events in these results reveal that errors come in many forms, so can we identify a taxonomy of error? Can we determine the ways in which some types might be stylistically useful?

9.2 The importance of visibility

> **Question 2:** What is the impact of visible risk on audience perception of DMI performance?

Study 3 aimed to investigate the importance of visible risk through performances using Keppi. Keppi is an instrument made in three versions, each with a degree of disfluent behaviour, and the risk state — a state in which error is likely — was displayed to the audience through lights on the outside housing. (See Chapter 6 for a discussion of the design of Keppi.)

This study was motivated by the story of Keith Jarrett's performance of *The Köln Concert*, in which Jarrett performed a legendary performance though he was confronted with disfluency in the form of a broken piano. As such, I wanted to test if risk, in the form of various degrees of disfluent instrument behaviour, would impact audience perception.

The post-hoc quantitative audience ratings of Enjoyment do not show that increased disfluency impacted audience enjoyment, and there was also no clear preference for performances with a particular level of disfluent behaviour. This lack of conclusive evidence may be due to study design, and perhaps an instrument that incorporated elements of chance in its behaviour, thereby being less predictable by the performer, may have demonstrated this effect more conclusively. Of course, it is also possible that the differences in instrument disfluency simply didn't matter to the audience, or may not register as salient. Further study is needed to gather conclusive data.

However, the multiple data sets gathered by the combined methodology was a distinct advantage in this case. By gathering not only quantitative ratings but also qualitative and real-time data that could allow the video to be coded at specific points, another phenomena emerged that affected audience perception: Visible *skill*.

Though all three studies involved experienced performers playing live, 'skill' was not a theme in the qualitative data from either Study 1 or Study 2. In the post-hoc qualitative data corpus for Study 3, however, skill was a theme that appeared often. 'Skill' was mentioned in the qualitative data in response to the question 'What did you like about the performance?' for all performers, except Performer C. In this case, skill (or more accurately, a perceived lack thereof) was a theme of responses to the question 'What did you dislike?'

A useful definition of skill comes from Fyans and Gurevich, who suggest that skill is an embodied quality that is a combination of control and effort [72]. Across the real-time data for all performers, the audience tended to identify areas where the performer was in control as ‘enjoyment’ events, and moments where the performer seemed to lack control as ‘error’ events. Further, the qualitative data for Performer C indicated that the audience not only perceived them to be lacking in skill, but specifically in effort.

Further, though there was no significant preference for any single performance nor any instrument behaviour type in the quantitative data, Performer C was rated significantly lower in both these quantitative aspects in both ratings of Enjoyment and rank ordering. In this way, it seems that a perceived *lack* of control and effort — or skill — had a negative effect on audience perception.

This aligns with Kruger et al.’s notion of the effort heuristic [122]. If Performer C was judged to be exerting less effort, then the effort heuristic states that their output would be perceived to be lower quality. Though I hesitate to draw any line between ‘enjoyment’ and ‘quality’ (both of these are extremely loaded words and their precise meaning and interaction in this context is outside the scope of this research), the connection between visible effort and perceived skill is certainly salient, and worthy of further examination.

The results of Study 3 indicate that visible skill is important to audiences, but care is necessary when drawing implications. I do not suggest that simply showing off for an audience is the only reasonable performance strategy, and I do not suggest that skill that is perceivable by the audience is the only skill that matters, or is the only skill that audiences appreciate. Of course, all skilful musical outcomes are underpinned by years of training and embodied knowledge that remain invisible. That said, the indication that visible skill is important to audiences implies that DMI designers may wish to consider how this could impact the physicality and behaviour of instruments and the interactions used to play them. Potentially, more robust implications for DMI design may result from further study into the embodied-control-and-effort model of skill, how this type of skill is communicated by performers, and how it is understood by audiences.

Though there is still much to learn, it is useful to highlight that the embodied-control-and-effort model, and the audience’s perception of such, moves skill into the physical realm. The results of Study 2 (which found that the audience enjoyment was affected by larger gestures used to play the larger version of MOAI) lend dimension here; in that study, I found that the physicality of the instrument itself has a part to play. Studies 2 and 3, then, point towards both the DMI's physicality, as well as its behaviour, having impact on the audience, but understanding if and how the fixed qualities of an instrument may intersect with its time-behavioural characteristics needs further examination.

A conclusion that can be drawn here is that focusing on the visibility of skill, and not the establishment of a musical frame of reference, may be a way for DMI practitioners to retain the radically experimental characteristics of DMI performance practice while still effectively engaging audiences. Dobrian et al. [62] suggested that better communication and expression can be achieved in DMI performance by establishing a common frame of reference, by fostering a community of players and audiences, as well as conventions of performance and repertoire. However, in Chapter 2 I presented at length a historical view of DMI performance practice, and highlighted how it continues to be radical and technology-led. In this way, a solution that imposes limits on an art form defined by its lack of boundaries seems incompatible. Though this incompatibility is in no way fixed, and an established frame of reference may well be useful in the future, focusing on visible skill may be an effective way to establish audience connection while still preserving the radical characteristics of this musical tradition that have made it a vibrant place of artistic innovation and progress for nearly a hundred years.

Finally, the findings related to this question suggest that risk may have nuance that is not fully explored. Though Study 3 shed light on the importance of visible skill, it is vital not to simply abandon the notion of risk, and rather to question what it means in relation to the wider findings. Risk is related to error, and when we consider the nature of error that relate to RQ1, the definition of risk begins to change shape: If risk is a state in which errors become more likely, then risk is only as meaningful as the errors themselves. As discussed in Section 9.1, this body of research found that error did not negatively impact audience perception, and Study 1 demonstrated that even an error-ful performance may still be greatly enjoyed. Where, then, does this leave risk in a DMI performance context? Perhaps the question has shifted from whether error is possible, to whether risk is possible. Indeed, this is an area with intriguing questions that may provide more specific insight.

9.3 The influence of the DMI's physical design

> **Question 3:** Can the physical design of a DMI affect the performative outcomes?

Both Study 2 and Study 3 offer insight into this question. First, as discussed in the previous section, the behaviour of the DMI affects performance in ways that are perceivable by the audience by demonstrating (or failing to demonstrate) skill. Secondly, Studies 2 and 3 found evidence that player choices were affected by the design of the DMI, and that these effects were perceivable by the audience (in the case of Study 2 this design was physical, and in Study 3 this was behavioural).

The physical scale of the DMI used in Study 2, MOAI, produced two different scales of gesture, and these differences were perceivable by the audience and made a significant difference to ratings of Enjoyment. This result confirms Jack et al.'s assertion [103] that the physical design of a DMI affects player decisions. The results here extend this finding, demonstrating that the physicality not only affects the player, but that they produce outcomes perceivable by the audience.

Of course, Study 2 tested only two versions of MOAI, and the results do not give any clear indication of how big a DMI needs to be to elicit more visible gestures, or at which point the smallness of an interface may obscure gestures from view. But, by comparing performances on two DMIs that were identical in hardware, software, proportion and interaction and differed only in terms of scale, this does illustrate that the differences in audience perception were due to the scale aspect of the physical design.

It is here that the notion of visibility explored in the previous section again comes into play, and it is here that more study is needed. For example, could we determine exactly how big a gesture needs to be to become perceivable? Further, what is the contribution of the performer, and the contribution of the instrument in producing playing gestures? Are there interventions in design features (as done with MOAI) or instrument behaviours (as done with Keppi) that might change the balance of this player/instrument relationship?

In Study 3, 6 copies of one instrument were produced, 2 of each study condition. All versions were physically identical, but differed in terms of their type of disfluent behaviour. The design value of interactive pluralism applied to the design of this instrument, as described in Chapter 6, appears to have been effective, evidenced by the wide range of ways that the performers played the instrument. Additionally, there is evidence that the instrument's behaviour affected the performers' choices. Evidence for this lies in the qualitative audience responses to the performances featuring a disfluent behaviour, which tended to focus on the physical aspects of the performance. This qualitative evidence was less prominent in the qualitative data corpus for the control performances. These results indicate that the findings by Jack et al. [103] are relevant not only to physical features of design, but may also extend to behaviours that present affordances in real time.

The implication for DMI performers and designers is that these aspects of visibility, particularly for the embodied-control-and-effort model of skill, may be addressable through the physical and behavioural design of the DMI. It is possible that this is due to how these factors affect performer choices in real time — this was found in the informal feedback from the Study 3 performers, but as the performer's perspective is largely outside the scope of this thesis in favour of emphasising the audience perspective, no precise conclusions can be drawn, though this remains an exciting area for future study. Nevertheless, the results relating to RQ3 indicate that the physicality of the DMI and its behaviour affect the performer, and have perceivable outcomes for the audience.

9.4 Combining post-hoc and real-time data

> **Question 4:** How can both post-hoc and real-time data be used to understand audience value judgements of a live music experience?

Throughout this body of work I experimented with ways to gain useful insights through the combination of real-time and post-hoc data. This section summarises the key methods used, their effectiveness, and their impact on the results.

9.4.1 Correlating button taps and ratings

In this body of work I have combined the real-time and post-hoc data in two ways. The first is using the post-hoc quantitative ratings of Enjoyment, and correlating these with rates of both 'enjoyment' and 'error' button taps per minute. The intuition was that, if post-hoc ratings are related to real-time behaviours, the rate of 'enjoyment' taps would be positively correlated with ratings of Enjoyment. Though this was sometimes the case and that correlation was statistically significant (as it was for one performance in Study 2), there was no consistent strong correlation, and most of these were not statistically significant.

This indicates that our opinions in real-time and our opinions just minutes later may be different, but also adds the insight that time itself may be an important factor in how we experience music. Simply being asked at a point further in time from the experience seems to change this opinion considerably, but these results don't provide a precise reason for this. Time itself may be the most influential factor, or it may instead be that, when asked to translate our in-the-moment opinions onto a numerical scale from 1-5 or distil them into descriptive language, additional cognitive processes may be engaged that shift our view, or perhaps time simply allows us to juxtapose our real-time experience with our memories, our taste, and our own cultural frame of reference. Whatever the reason, future enquiry that probes how post-hoc and real-time perceptions of music differ has the potential to produce intriguing results.

9.4.2 Individual opinion

The amount of data collected in Study 1 allowed for an explanation of opinion and real-time behaviour to be examined for individuals, and for some patterns to emerge. Because Metrix created usernames that participants wrote on survey books, there was the potential to view the real-time behaviours of individual participants and compare this with their post-hoc opinions. In Study 1 there were a small number of users who were particularly prolific in their button use that made this examination useful.

In Study 1 there there were indications of incongruous real-time behaviours and post-hoc opinions, lending further evidence for profound differences between opinions in real time and those collected after the fact.

This is not to suggest that one opinion is 'truer' than the other; rather, it demonstrates that we report our opinions differently depending on when and how we are asked. If I had asked different questions in Studies 2 and 3 I may have found a way to explore this further on an individual level, but even in viewing the audience as a whole, both sets of data yielded important insights that the other could not provide — the qualitative data, for instance, had rich context that the real-time data could not have, and the real-time data had time precision that is absent in post-hoc methods — so both are valuable.

9.4.3 The influence of time

Finally, these results present questions about the effect of time on audience experience. There is no conclusive evidence that real-time button taps of either 'enjoyment' or 'error' are correlated with post-hoc Enjoyment ratings. The presence of error does not seem to impact enjoyment, but the incongruency between real-time indications of 'enjoyment' and the post-hoc ratings of Enjoyment remain unexplained. It is possible that audiences enjoy engrossing experiences, and when they are engrossed audiences are less likely to tap buttons. It is possible that an audience's assessment in the moment and their reflective opinion after the fact may engage very different parts of the mind. Time may have an effect on our perception, and perhaps distilling our opinions into language does also. This is a particularly rich area for further study of the perception of live audiences, and also indicates that approaches to studying these audiences beyond post-hoc surveys has useful applications in this area.

9.5 Reflections

In this section I reflect on the contributions of this thesis and the process of producing them, in order to highlight important lessons for future researchers in this space.

9.5.1 HCI's role in DMI research

The primary contribution of this thesis is new insights into the nature of error as experienced by audiences.

As DMI research started off (and, arguably, still remains) a curious corner of HCI, it is common for the DMI research community to look first to HCI for ways of understanding phenomena within DMI performance.

The notions of error that develop out of this research, however, find no home in HCI. Historically, error theory within HCI is task-based, developed in response to industrial tragedies caused by poor design. Considering an error in a musical performance with the same criteria as an error that, for example, caused a plane crash, is unlikely to yield any relevant insights.

The insights on error that emerged through this research — that it is nuanced, is not negative, and exists outside musical boundaries that typically take the shape of stylistic conventions — go beyond generalised ideas of error in HCI, and explore error not as it relates to computers, but as it relates to this specific community of musical practice.

Perhaps, then, HCI methods should not be the first place DMI researchers look for insight. This is not to say that HCI has nothing to teach us; particularly with the recent emergence of third-wave HCI, as discussed in Chapter 2, HCI research is pivoting towards understanding computer activity that is not task-based and is much more ambiguous in its goals.

9.5.2 Connecting culture and history to DMI research

Many, if not most, DMI researchers are also DMI practitioners. Because of this close coupling between those engaged in this musical practice and those studying it, I had an expectation at the start that I would be able to find a cohesive history of DMI practice from which to begin.

However, this does not exist in any real sense; there are pockets of research tracing the origins of instruments, for instance — Thor Magnusson's current work on organology is a particularly good example [133] — but there is an absence of overarching critical discourse. Perhaps this is precisely because those engaged in the practice are also the ones doing the research, and critical discourse might require a certain degree of distance. It might also simply be another case of DMI performance's radical nature resisting formalised description.

In some ways this struggle had very useful outcomes — the survey of contemporary DMI practice in Chapter 2, for example, provided me with a topography of this discipline, and a new appreciation for the huge range of performance practice that it includes. I came to see the label of 'DMI' as not a way to exclude forms of practice, rather a refreshingly flexible term that can stretch to around the huge diversity of musical practice within this community.

But, one researcher tracing a history or just drawing a line around a practice at one point in time is not critical discourse, and the historical context presented in this thesis is only a one viewpoint. DMI performance is now maturing to a point where we are increasingly interested in issues beyond the technical details of DMIs, and fostering a critical discourse can help us understand what we're doing when we make music with computers, and how it is different from other ways of making music. There is much to explore here, from the instruments and technical details to broader social structures such as the role of universities in this community, the impact of this research on culture at large, as well as perspectives on the DMI practice that include voices of women, people of colour, and queer folks. This research has taught me the importance of critical understanding, and though some of this development is already underway, both DMI research and DMI practice will benefit from a wider and deeper conversation.

9.5.3 Creating DMIs for experimental purposes

Building the instruments for experimental research had a number of advantages, primarily the problem of finding appropriate DMIs to use in studies. For the first study finding musicians who played appropriate instruments that could be played in two distinct musical styles was very difficult, and I realised that this research may be limited by the available instruments and players. Designing and building the instruments meant that I simply needed to find willing collaborators.

Secondly, control over every aspect of the instrument meant I could design the experimental axes of comparison. For Study 2 the axis of comparison was MOAI's proportions, and for Study 3 this was Keppi's degree of disfluency. Because I was controlling the design, I also I had control over the study question, instead of having to compromise my interests based on available DMIs and performers.

By directly comparing audience perception of DMIs based on specific variables, I was able to mitigate some of the confounding factors that often complicate this kind of research. In turn, this research could take place in a context that concentrated on preserving important multimodal factors of live performance.

But, simply designing any DMI with experimental variations is unlikely to be enough. The most important aspect of the DMI design process was that I wanted to produce instruments, not simplistic experimental tools. It is through engaging with instrument design that musicians will be adequately challenged and creatively engaged.

Finally, designing DMIs for experiments means that I have watched other musicians using both MOAI and Keppi, and I have had insight on their effects on audiences. This is valuable insight that DMI practitioners are rarely afforded, as we are often the only players of our instruments [144]. For example, Study 2 allowed me to watch MOAI from the audience's perspective, and see how it communicates on stage at large and small scale, and this has informed my choice of gestures and the stance that I use to play MOAI. Study 3 provided the opportunity to see the effects of Keppi with various behaviours in the hands of six experienced musicians. This rare vantage point made clear the next steps for Keppi: Instead of having a regular interval between ticks, for instance, incorporating an element of chance might keep this disfluent in the long run.

9.5.4 A new methodology for audience research

The combined methodology, in many ways, has been very successful. It has allowed not only a comparison of real-time behaviour and post-hoc data, but has also enabled precise analysis of the video documentation, leading to insights on how audiences perceive error and enjoyment in DMI performance.

Metrix, which started out as a little Node.js experiment, has been invaluable in this process. It allowed me to leverage available technology, easily collect data, control what was collected and how it was stored, process that data in whatever way was most useful. Using web technologies also meant that I could also easily iterate on the interface.

Of course, there is nothing about this methodology that has been rigorously tested or ratified — this thesis only presents its first three uses. Much still needs to be investigated, such as the best way to code video data for 'error' and 'enjoyment' events for instance. Nevertheless, this methodology has allowed me to make vital connections to audience opinion to moments in time, and allowed a peek into audience reactions in real time.

It is also a method that focuses primarily on audience opinion. There are a plurality of audience research methods involving physiological feedback, but as I detailed in Chapter 2, these affective properties are not necessarily indicative of subjective evaluation in this context. As such, this methodology does not query valence, and instead roots evaluation in subjective opinion. This allows the real-time and post-hoc data sets to be integrated, and presents an opportunity to use one to inform the other. Though real-time physiological data in conjunction with post-hoc evaluation may lend valuable insight into the valence of real-time physical affect, it is not subjective; Metrix, however, allowed me to combine real-time and post-hoc data which are both opinion-based.

Most importantly, this methodology has resulted in the contribution of a practical tool to the DMI research community, which currently still leaves the audience very much under-studied. My hope is that the availability of Metrix and this way of using it will lower the barrier to entry for DMI researchers, and more audience-focused research will take place, yielding more insights with new nuances.

9.5.5 Doing research in a musical context

Studying the audience in a concert context made me keenly aware of the contextual factors affecting how a performance is perceived. Every aspect of the musical setting — which I came to think of as an 'ecology' of factors — is a potential source of contextual data: The lighting, the people, the seats, the flow of the show.

While this fidelity to the concert 'ecology' is a major strength of this research, achieving it is an extremely complex process; at most, I have only scratched the surface of this kind of investigation. In future, however, it would certainly be interesting to extend this approach into other contexts where DMI performance happens, from the night club to the festival to the website and beyond, to understand not only how audience perception differs in these contexts, but how DMI music adapts to the setting in which it is performed, and the role that the setting has to play in the live music experience.

9.6 Summary

In this chapter I summarise the findings from each study as they relate to the larger questions guiding this research. There are a number of implications for DMI performers and researchers, some of which may extend to HCI research as well. Further, there are a number of potential areas for future research that have been illuminated through this work. Additionally, I review the four contributions of this work, and provide reflections around these.

In the next and final chapter I add final reflections on the wider implications of this research, and concluding remarks.

Chapter 10

Concluding Remarks

Feci quod potui; faciant meliora potentes.[1] [46, quoted]

It is a curious thing to study music performance as if it is a fixed and knowable quantity that gives way to scientific examination. In trying to resolve these two worlds, I have gained significant insight not only into how we make art and what it means to audiences, but also into how the methods, goals and values of scientific research and musical performance resist one another. Though they may be curious bedfellows, the process of science and the process of music performance are not irreconcilable. I close this thesis with thoughts on instruments, audiences, and error gained by doing this PhD, as a message in a bottle tossed out to sea for those scholars who may continue this kind of work in the future.

10.1 On instruments as tools of creation

The most dangerous phrase in the language is, 'We've always done it this way.' — Grace Hopper

Every artist has an intimate relationship with the instrument of their craft: A musical instrument, a favourite pencil, a well-worn chair, a good pair of pliers. Often these instruments have been adapted, extended, and repaired over time, and every artist has an intimate understanding their favourite instrument's strengths, shortcomings, and idiosyncrasies.

[1]'I have done what I can; let those who can do better.'

The prospect of understanding the relationship of a masterful player with their instrument of creation and transferring these qualities to computers is tantalising. But, though a good instrument in the hand of a master can make magic, it is important to remember that masters are not made through design interventions alone. This has been, perhaps, the most satisfying aspect of this study: That there is still no short cut, no digital intervention that can substitute for the moments of inspiration, the skill gained over decades, the leaps of imagination, or the desire for self-expression that can come only from years of experience of being a human and moving through the world.

This is not to say that the insights of this research are not useful, but rather that DMI research must perhaps become comfortable with dissatisfaction, or at least a lack of resolution. All musical instruments were, at one point or another, novel interfaces, but until just a few decades ago an instrument evolved slowly in response to audiences, players, material advances and musical trends.

It is important to recognise that as DMI researchers, performers and composeres, we have unique insight into how computers impact creative practice. Like all musicians in history, this community has been a ferocious early adopter of technology, and has been harnessing it to make instruments of creation for over 100 years. It is only relatively recently, as the digital dew has dispersed over human activity and is now sinking in deep, that the rest of human activity is following suit on a massive scale. This means that the concerns of DMI research are not confined to a curious corner of HCI scholarship where people make weird music. Instead, this community may have a monumental head start in insight and knowledge in this realm, and as such has the potential to lead the design of digital tools for creative applications.

We must recognise, however, that any digital instrument must do what humanity's best instruments, musical and otherwise, have always done. The paintbrushes, pencils, the musical instruments that need to be struck, plucked, blown and bowed endure because they are instruments of potential. They allow for continuous innovations in ways they can be used and what they can be used for, unlimited by specific task. They lend themselves to the exploration, innovation and progress that comes from doing things the way they're not meant to be done, by committing errors and observing the outcome.

Error, here, is perhaps an essential ingredient, because error encapsulates the human capacity for chaos. HCI has spent decades understanding human error and its destructive power, mitigating it and in the process making the modern world astonishingly safe.

But, this narrow understanding of error and view of it as negative is unhelpful in this realm. A good instrument will allow an artist to do it wrong, in order to discover a new way of doing it better. We must engage creative thinking and its process- and experience-based knowledge to discover the nature of useful deviations, how they come about, and how our future digital musical instruments, and other instruments of creative production, can usefully incorporate them.

10.2 On audiences

> *A nice blend of prediction and surprise seems to be at the heart of the best art.*
> — Wendy Carlos [4, quoted]

The audience is as old as music, and is a vital component of live performance. Every audience is also made of people, and therefore holds all the characteristics one might expect from a human being — contradictory, elusive, mercurial. Though the presence of an audience is so ancient that we take it for granted, the people who make it up are as multifaceted and complex as they have ever been, and this will not change.

There are still some commonalities. One is that audiences, in typical contradictory fashion, respond to two things: Newness, and pattern. Too much of one, and the other is lost. Where the optimal balance is between them, and where one stops and the other begins, is perhaps the most tantalising challenge in this domain, and it is in the understanding of this point of flux that art is equipped to succeed, but science can gain understanding in trying to observe.

Audiences are often viewed as inconvenient or problematic sources of data in DMI performance research. Audiences are challenging, definitely; they are ever in flux. But, it is only through understanding the perspective of audiences that we can understand how computer-based work can communicates to that part of us that music is for. It is in the audience — the real, live, difficult, inconsistent, contradictory audience — where there are indications of how our new digital instruments should function.

What I have come to understand about audiences through this work is not about their perception or their difficulty. Rather, it is that they want to like things. They want to understand. In studying opinions in real time, I consistently got the sense of the audience *trying* — to find patterns, to connect, to appreciate. Audiences, in this way, are not just an inconvenient necessity for this kind of work, but rather the very essence of our human need to connect, and a beautiful expression of the best generosity, openness, acceptance and respect that strangers have to offer one another.

Appendix A

Required statement of originality for inclusion in research degree theses

I, Sarah Marie Astrid Bin, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

Appendix B

Metrix: Heuristic evaluation workshop results

Username	Device	Expert?	1 - severity	1 - comment	2 - severity	2 - comment	3 - severity	3 - comment	4 - severity	4 - comment	5 - severity	5 - comment	6 - severity	6 - comment	7 - severity	7 - comment	8 - severity	8 - comment	9 - severity	9 - comment	10 - severity	10 - comment
explainRound		Yes	2		4		3		3		2		2		4		3		0		0	
		No	2	Sometimes not sure if the buttons are being pressed. button-down state doesn't always show.	0	Cross tends to mean error and smiley face tends to mean like. This seems appropriate as smiles = approval and cross = something is wrong.	3	Cannot correct yourself if you press the wrong button. Perhaps avve bar at the top with most recent decisions	0	Cross means error in lots of UIs and smiley face means like in lots of UIs	0		0		0		0		0		0	
purposeCrop		Yes	0		0		4		2		4		4		4		0		4		1	
somewhereSpeech		No	1	Could be more feedback when you press	0	symbols work.	3	No back button or undo press function	0		0	Can't see any errors that could happen	0	Very visible	0	Not needed	0	Very simple interface	3	No undo	0	No documentation needed
becausePolitical		No	3	Not clear whether I pressed the button.	0	Yes, :) and X are obvious to me	3	Not clear if I can change my choice	1	No obvious feedback to [?]	0	No errors possible	0	only 2 options! No problem	0	n/a	0	simple UI	0	No errors can be made?	2	There's no docuemntation to be found but it's not really needed
toyAlready		No	0		0		4	No way of changing input	0		0	n/a, no errors occurred	0		2	Doesn't automatically turn to landscape	0		0	I don't think you can make an error but not sure	0	
ourselvesAcres		No	0	If the app gave feedback it would interrupt the performance.	1	The two symbols don't match, they could be :) and :(or tick and X	0	I assume it's for a large audience and a mistake could be tolerated	0		0		0		0		0		0		0	
		No	1		0		3		1		0		0		0		0		1		2	
payMotor	iPhone 6S	Yes	1		0		1		0		1		0		1		2	Fixed apps in the system [ed - what the hell does that even mean]	2		3	
guessThick	iPhone 6S	Yes	3		1		4		0		1		0		1		3		2		0	
affectEntirely	Android, XiaoMi		0		1	The :) and the X are not exact opposites. Maybe they should be :) and :(	0		0		0		0		0		0		0		2	A link to a youtube video or about page could help.
someFrozen	iPhone 6		2		1		1		2		2		1		1		0		3		3	
serviceRate	iPhone 5	Yes	0		0		2		1		1		0		1		1		2		2	
needleFree		No	0		0		3	Can't quickly undo	0		0		1	Can't remember what button last pressed although doesn't matter but would like extra confirmation that pressed right button but that might end up more confusing.	0		0		3	Can't quickly undo	0	
butSyllable		Yes	4		0	Yes	3	No	0		0		0		0		0		0		0	
sureTwice		No	3		1		4		1		0		1		3		1		0		1	
eraserVolume	iPhone 5S	Yes	3		0		3		1		2		0		0		0		0		0	
comfortable Finally			0		1	Why :) and X and not :(	3	Can't see how to do this	0	seem to	0	Human error still possible	0	Yes, from symbols	2	Maybe tricky if you want an in-between	0		0	Not sure	0	No
shoutOffer		Yes	0		0		2	No choice for mistakes! Could have a confirmation in group selection screen.	1	Red button on Group 1 can change	2	Error when selecting groups isn't recoverable	0		0		1	Again, colour of red group	2	Group selection	1	Maybe short text at the beginning in case someone missed the video?
AVERAGE SCORES OF SEVERITY:			1.32		0.53		2.58		0.68		0.79		0.47		1.00		0.58		1.16		0.89	
AVERAGE SCORES FROM EXPERT:			1.63		0.63		2.75		1.00		1.63		0.75		1.38		1.25		1.50		0.88	
AVERAGE SCORES FROM NON EXPERTS:			1.25		0.25		2.88		0.38		0.00		0.25		0.63		0.13		0.88		0.63	

Appendix C

Study 1: Audience questionnaire

PERFORMANCE 1: DIANNE VERDONK

Username: ____

1. **How much did you enjoy the performance?**

1	2	3	4	5
Not at all				Very much

2. **How interesting did you find the performance?**

1	2	3	4	5
Not at all				Very much

3. **Did you understand how the instrument worked?**

1	2	3	4	5
Not at all				Very much

4. **What did you like about the performance?**

____

____

____

5. **What did you dislike about the performance?**

____

____

____

6. **How would you describe the performance to a friend?**

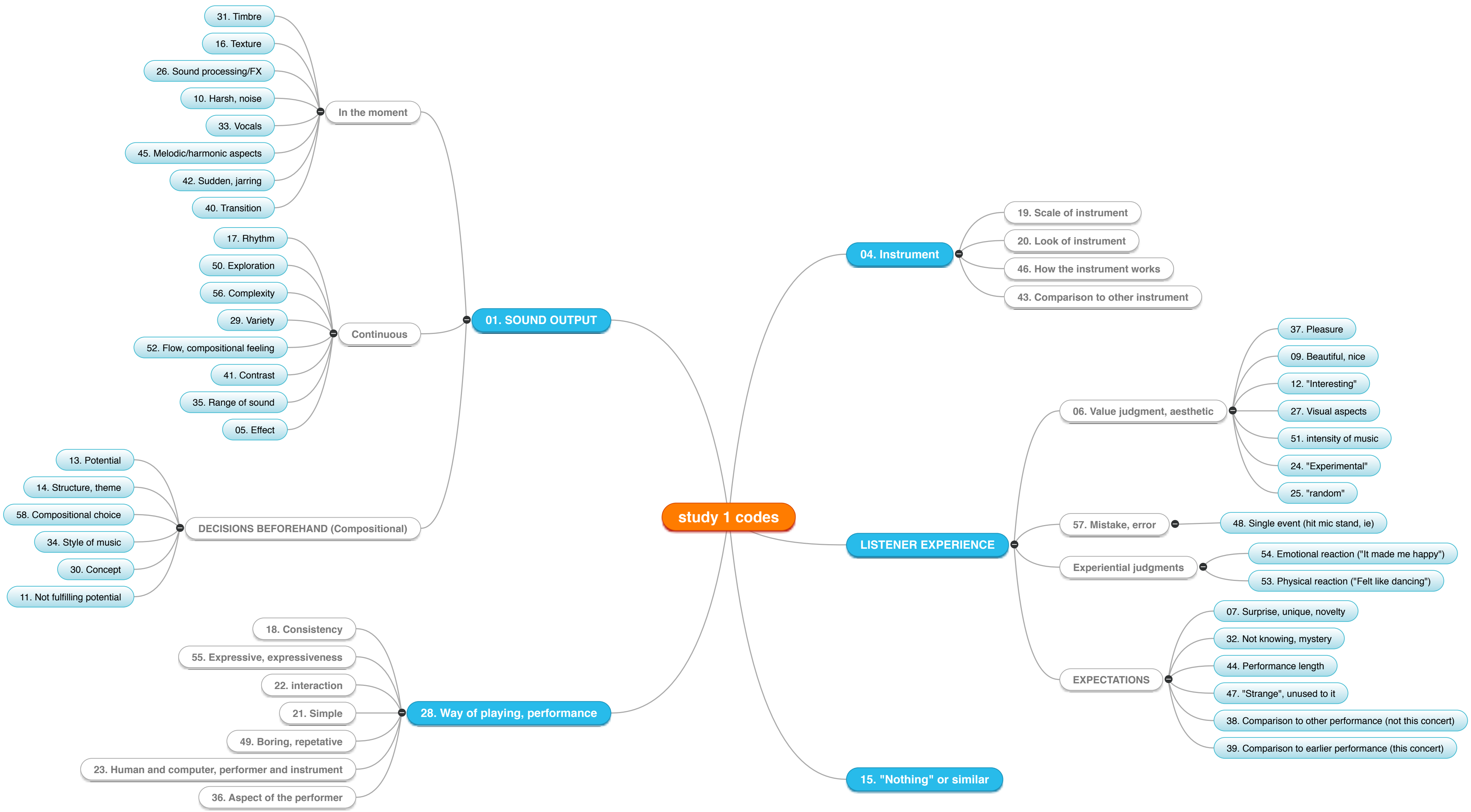
____

____

____

Appendix D

Study 1: Qualitative analysis



code_map_s01 I disastrid I 2017-11-04

User	Group	P1_LIKE	P1_DISLIKE	P2_LIKE	P2_DISLIKE	P3_LIKE	P3_DISLIKE	P4_LIKE	P4_DISLIKE
1	1			4 13 1	1 14	33 14		52 53	
2	1	4 6	15	3 12 36	2	4	33	34 28	15
3	1	1 16 7 12 20	17 14 18	12 16 19	1 10	28 9	1 4 49 28	6 28	
4	1	22 21 23 9 1		24	10	54	1 4 21	54 9	48
5	1	9 4 22 23	25 14	1 26 16	10 1 27	1 21 33	42	6 17 26 28	15
6	1	20	11	29 1	30	23	23	23 1 28	29 A
7	1	4	27 36	1	26	33	15	28	15
8	1	1 7 31 32	13	1 28 31	14	52	28 57	39	52 56
9	1	20 7	9 33	15	34	14	28	45 28	42
10	1	22 4	28 33	23 28	1 10	39	28	28	15
11	1	35 36	14	36 29 13	17	33 34	17 57	6 54	28
12	1	4 1	29	36	14	23 33	57	17	14
13	1	1 20 23 36	10 37	1 12 29 38	40 10	39	57	39 1 52	56 14
14	1	33 6	15	14	14 10	28	15	38 14	54
15	1	9 23 41		29 1 9 6	42	23 9	58	53 29 1	15
16	1	7	43 1 6	7	6 10	33	47	39	10 47
17	1	33 38	28	1	17	54	34 6	50	15
18	1	23	44	29 28	14	9 28 58	17	28 53 6	17 14
19	1								
20	1	12 1 29 33	45	17 7	10 1 42	6 45 33		17 45 6	56 58
21	1	20 28	1 6	20 28		53 33	57 1	17 40	15
22	1		17	1	10	33 9	23	28	
23	1	21 29 4 12	45	4	46 4	4	57	6 28	4 32
24	1	23		23		33	28	28 23	
25	1								
26	1	23 27 28	27	31	25	33 28	49	17 39	27
27	1	7 4 23 33	4	6 28	42 10	39 58	47 1 57	58	40 57
28	1	6 1	33	6 1		28 4		53 17	
29	1	28 7	44	29 1	14 38	28 4		17 6	
30	1	29 4 1		14 25	14 25	9 33 54	58	17	15
31	1	28 4 7	17 A	26	28	33		53	
32	1	4 7	34 38	17	17 A	23 14	15	28	48
33	1	6		6		28	10	56 5	
34	1	20	10	20	1 47	33 9		34 6	
35	2	1 29	1 29 A	1 12	10 1 14 26	17 33 6	57 1	58 39	1
36	2	26 29 28	15	4 12	28 17 A	7 23	57	28 17	27 32
37	2	1	48	15	10	23		17 1	49
38	2	21	28 29 B	28	28	58 23		50	44
39	2	20	14	17	10	23 33	15	6 26	10
40	2	23 4	5		10 42	23	49 55 A	54	
41	2	7 28	1 49 29 A	17 7 29 1	10 42	33	57	7 17 28	14 40
42	2	28 4	14 17 A	17 39	14	17 31 58	57	17 14 28	51 48
43	2	23	44	1	11	4 23	58	58 17 6	28
44	2	33	1	50 1	10	7 33 4 31		50 17 40	
45	2	7 1	14	1 7	14	54 28	57	6 17 29	15
46	2	28 29	1	29 1	49	29 1 28 23 41		58	57
47	2	28 7	5	36	49	33 23 28	57	17 6 28	32
48	2	23	31	1 26	14	6 4 7	15	28 56	48 17 A
49	2	6 7 4	7	1 29	32 4	23 4 7 28	1	12 1 28	32 30
50	2	7 28	1	28	10	15 28 58 55 A	4 58	28 17 56	10
51	2	4 7	6		14 10	9 23		28	56
52	2	4	14	1	10	58	49	53 17 6	
53	2	4 28	1	51	42	54 6 9 23 33	15	6 53	10
54	2								
55	2	26		26 30		33		33 17 34 58 6	
56	2	33			10	23		17 29	
57	2	1 28	29	36 30	10	33 23 39 45	30	28 17 54	49
58	2	33	49	7 30	14	17 45	1 49	54 34	1 6
59	2	1 29	1	7 1	44	33 23	57	17	15
60	2	28	10		10 14	9 36	28	6 17 29	58
61	2	28 1 29	33	1 28	10 14	23	23	17 29 7 1	14
62	2	4 7	34	36 28	10	23	28	17 58 28	15
63	2	7	34	7 1 14	10	6 23 41	15	52	42 10
64	2	7 4	45 1	31	10	33 9	35 4	6 17	15

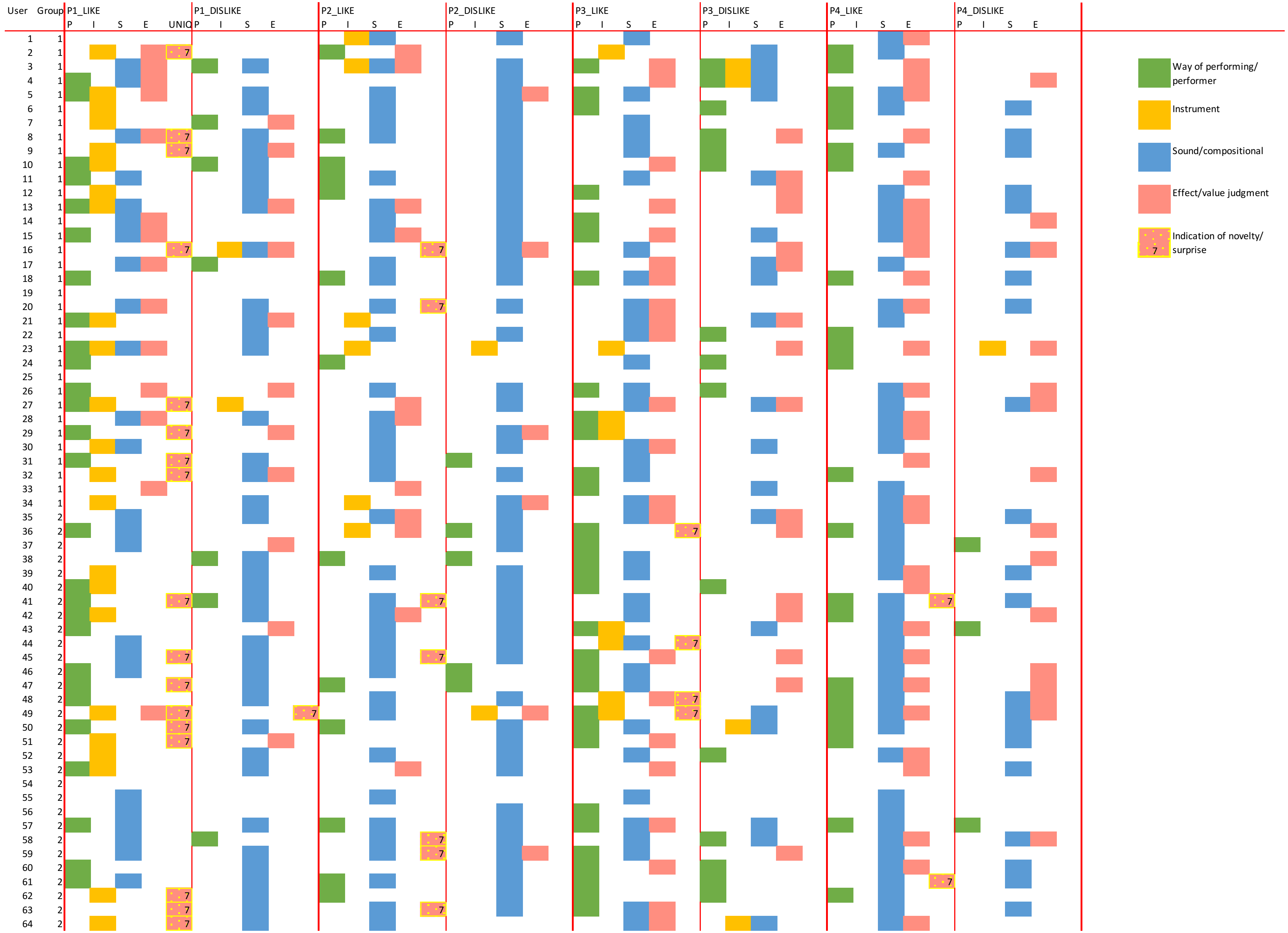
4 INSTRUMENT
19
20
43
46

15 NON-ANSWER

28 WAY OF PLAYING
18
21
22
55
49
23
36

1 SOUND OUTPUT
IN THE MOMENT:
10
16
26
31
33
40
42
45
CONTINUOUS:
17
50
56
29
52
41
35
5
25
COMPOSITIONAL DECISIONS:
13
14
58
34
30
11

LISTENER EXPERIENCE
6 VALUE JUDGMENT, AESTHETIC
37
9
12
27
51
57 MISTAKE, ERROR
48
EXPERIENTIAL JUDGMENTS
53
54
EXPECTATIONS
7 UNIQUE, NOVELTY
32
44
47
38
39



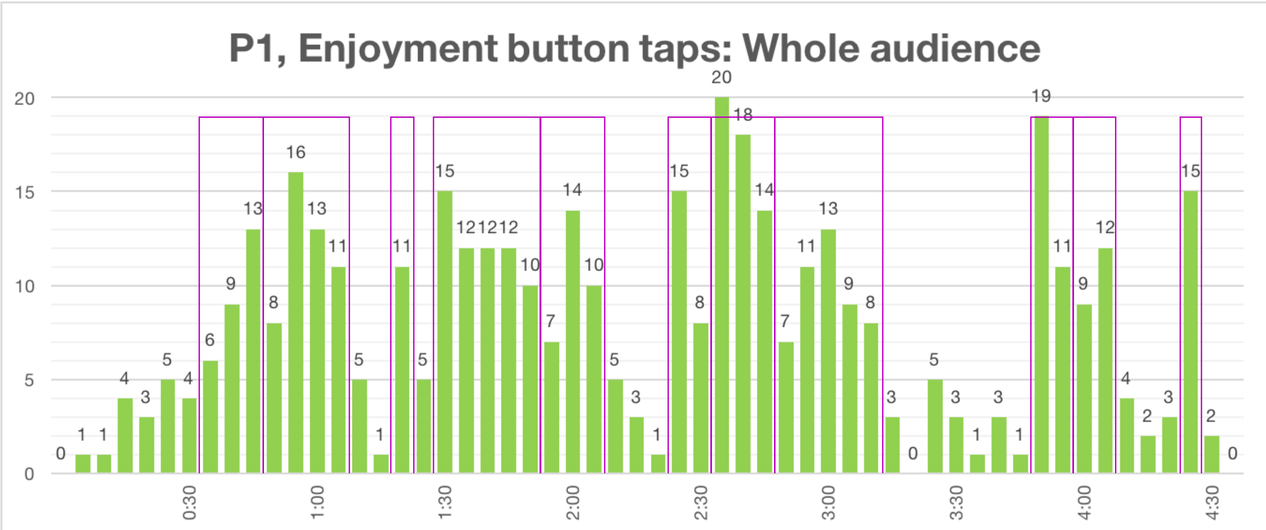
User
Group
P1_LIKE
P I S E UNIQ
P1_DISLIKE
P I S E
P2_LIKE
P I S E
P2_DISLIKE
P I S E
P3_LIKE
P I S E
P3_DISLIKE
P I S E
P4_LIKE
P I S E
P4_DISLIKE
P I S E
Way of performing/ performer
Instrument
Sound/compositional
Effect/value judgment
Indication of novelty/ surprise

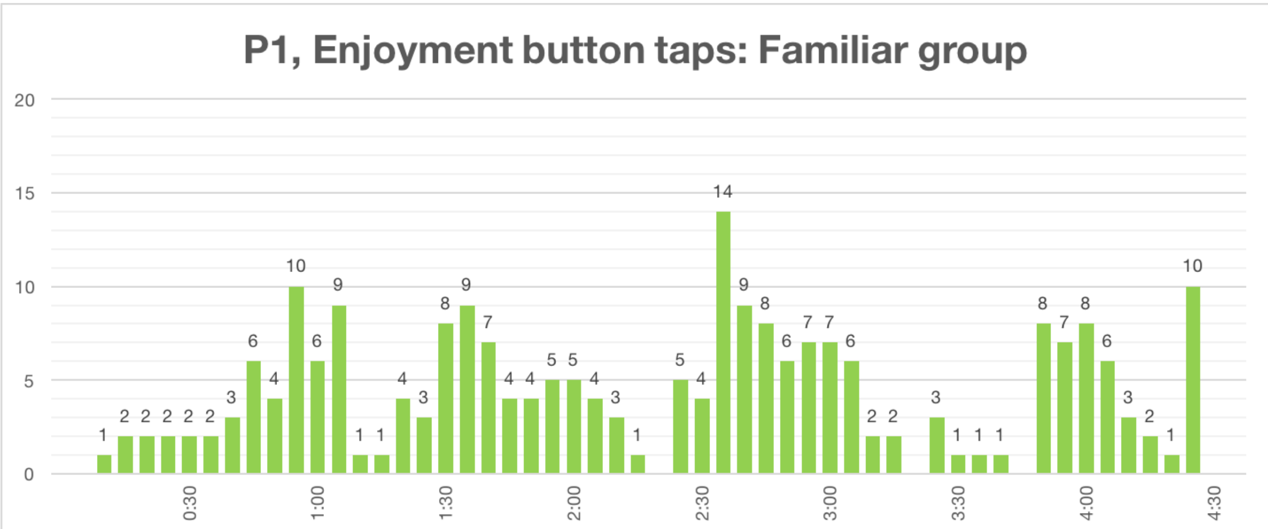
Appendix E

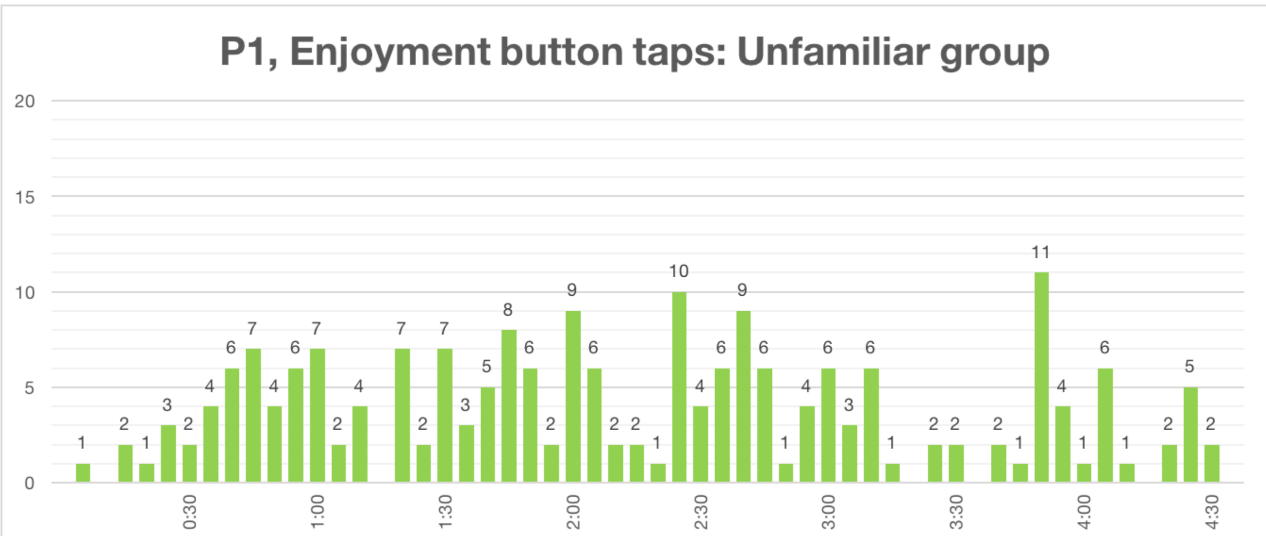
Study 1: Real-time analysis

Performance 1: Enjoyment
Real-time data visualisation

Dianne, experimental

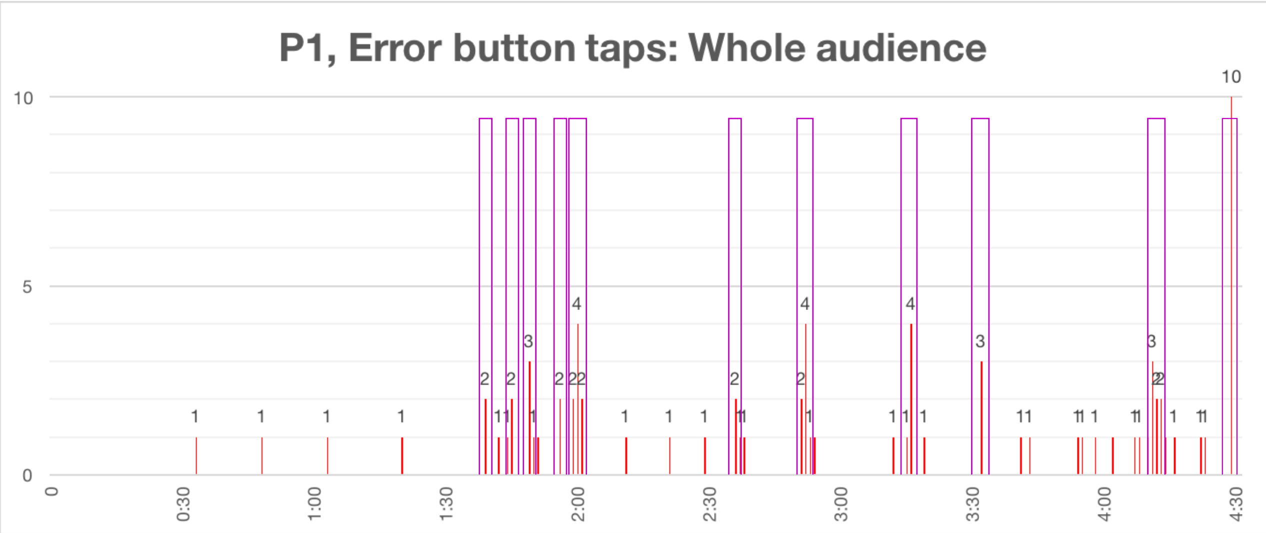


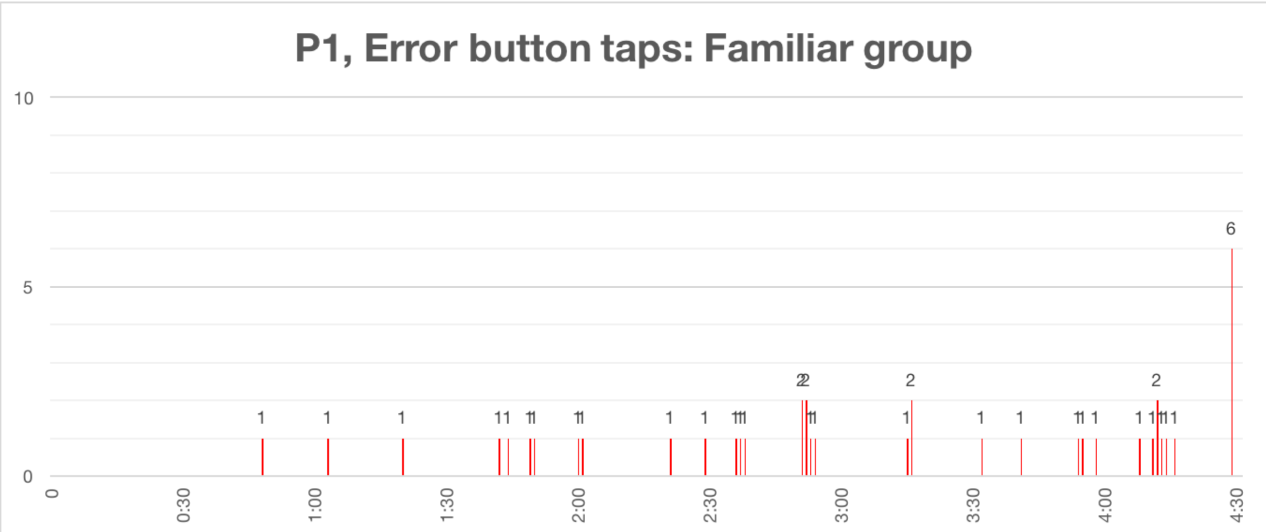


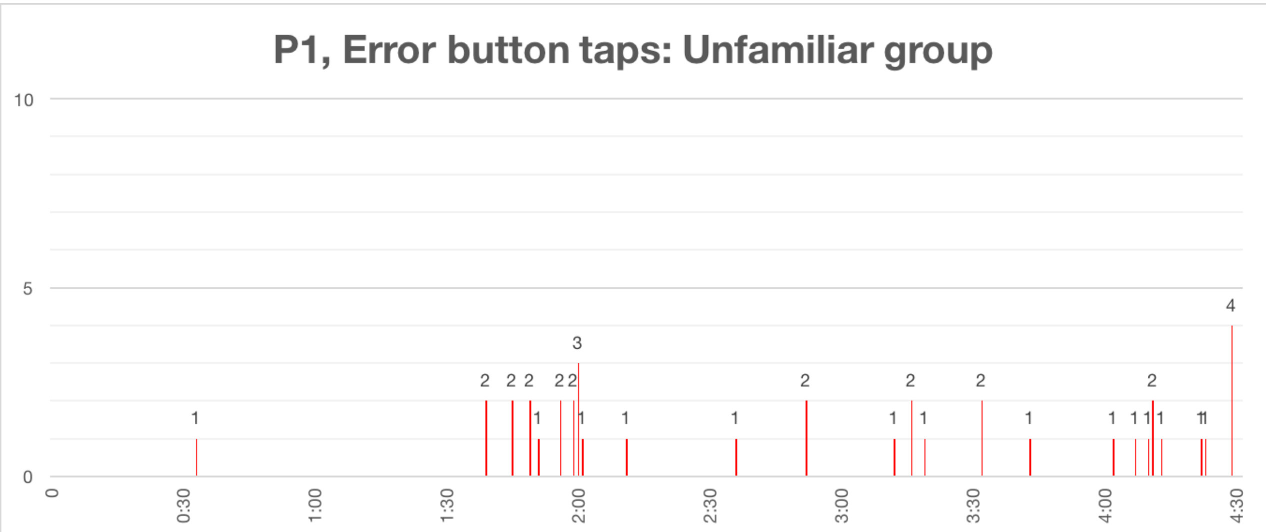


Performance 1: Error
Real-time data visualisation

Dianne, experimental

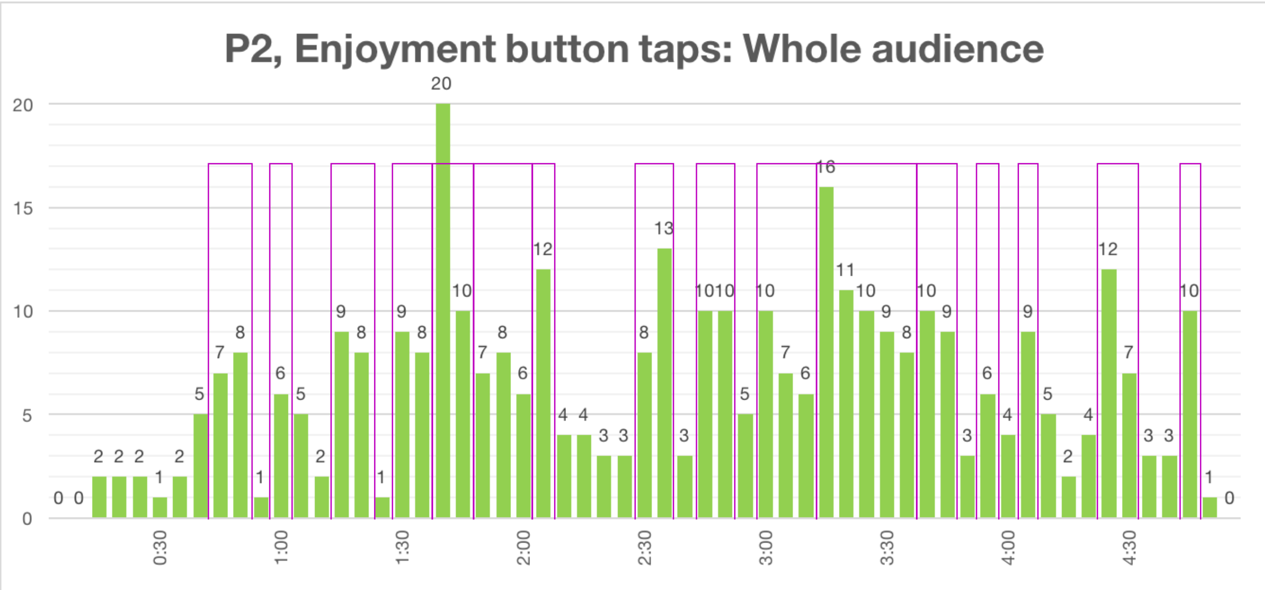


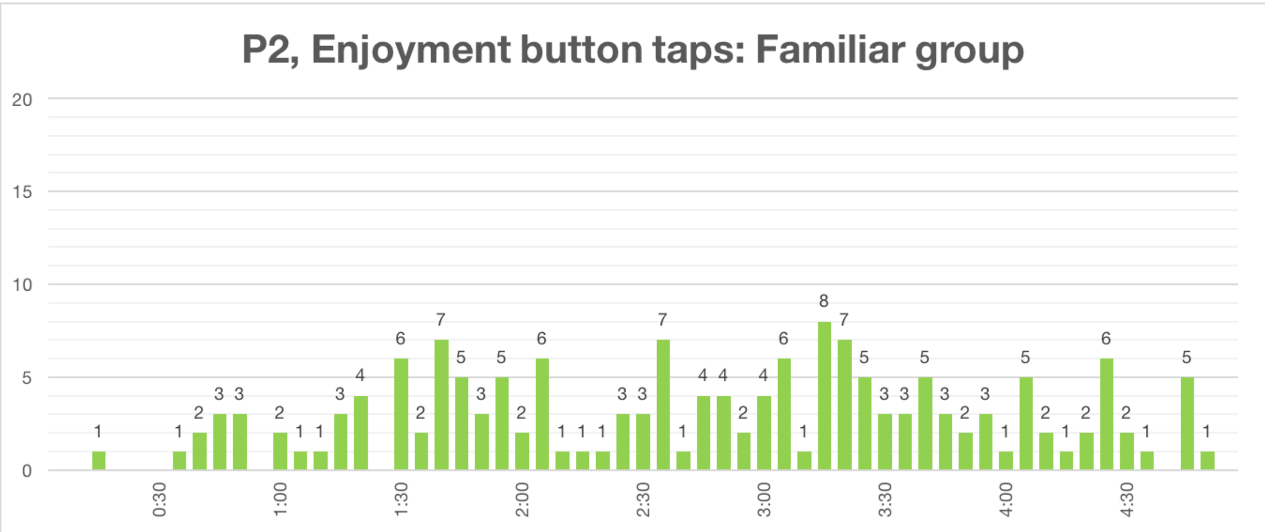


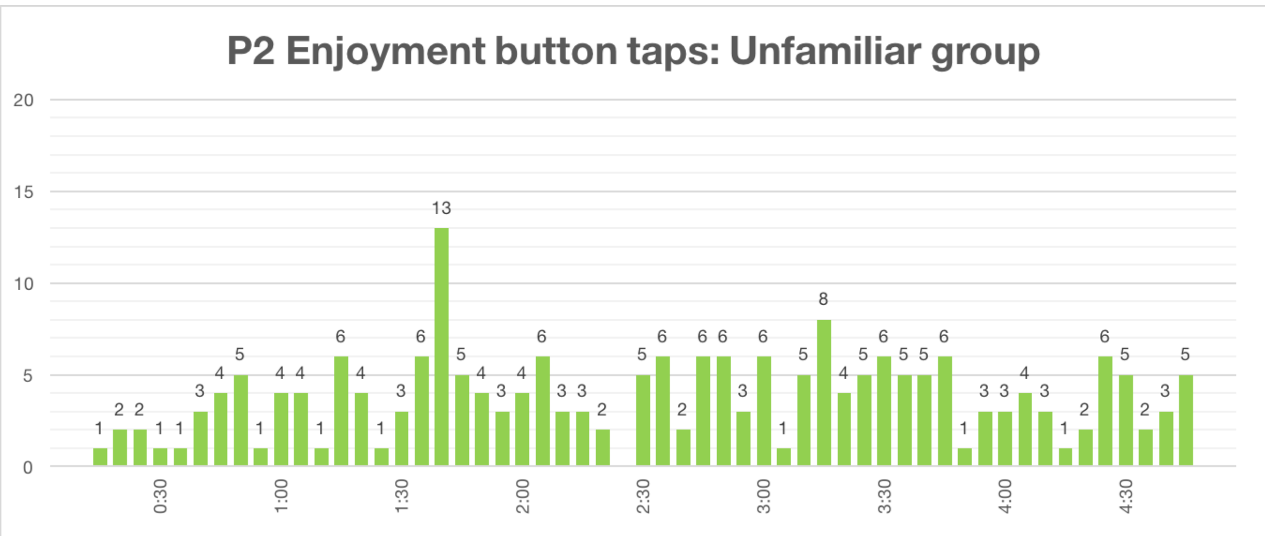


Performance 2: Enjoyment
Real-time data visualisation

Tim, experimental

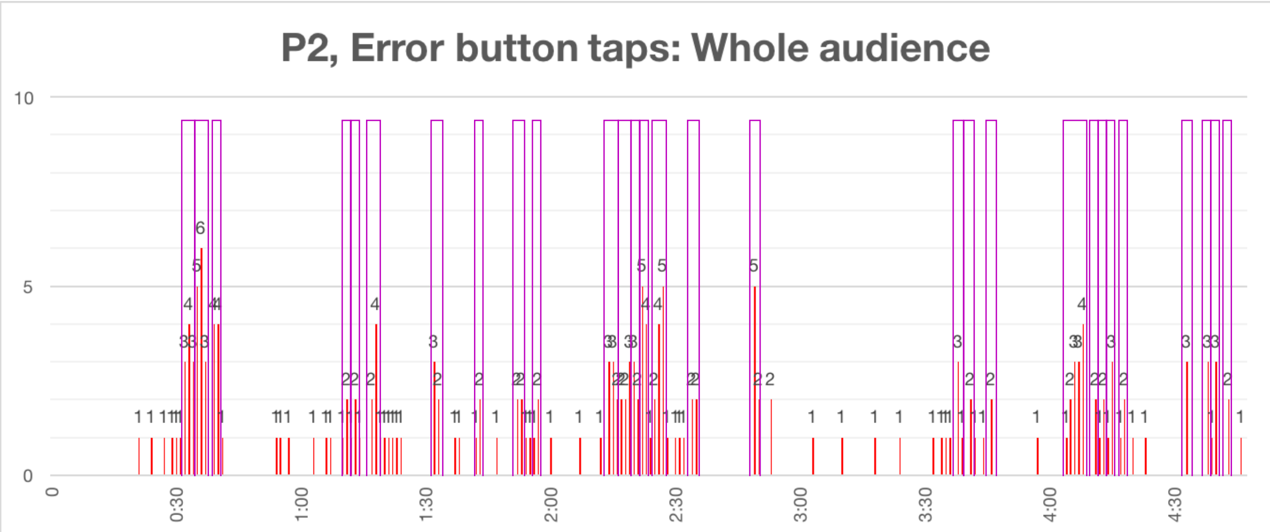


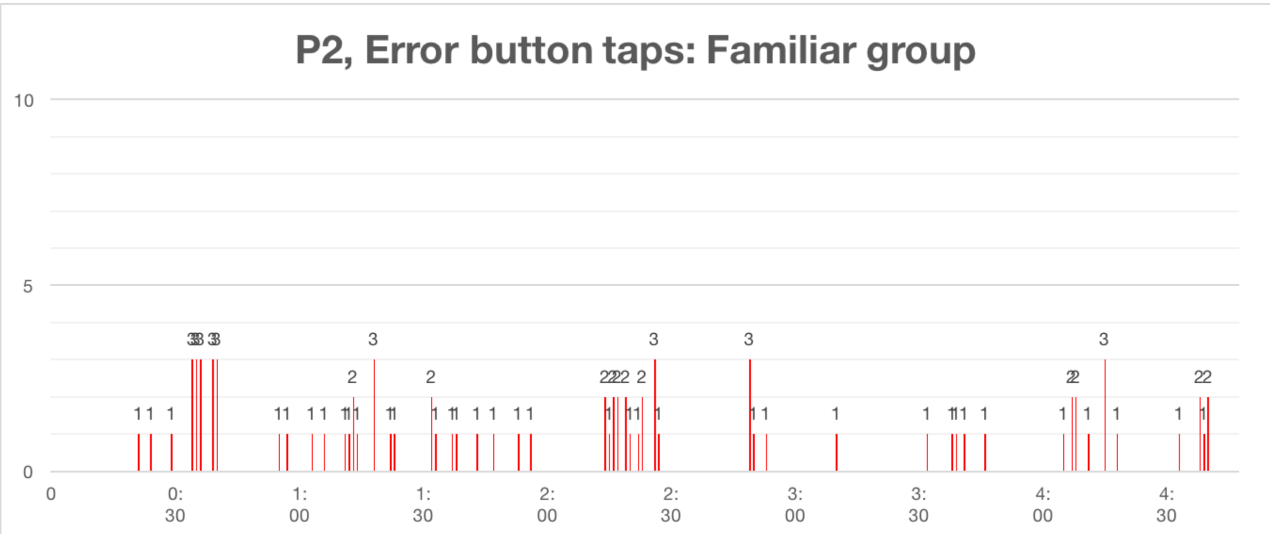


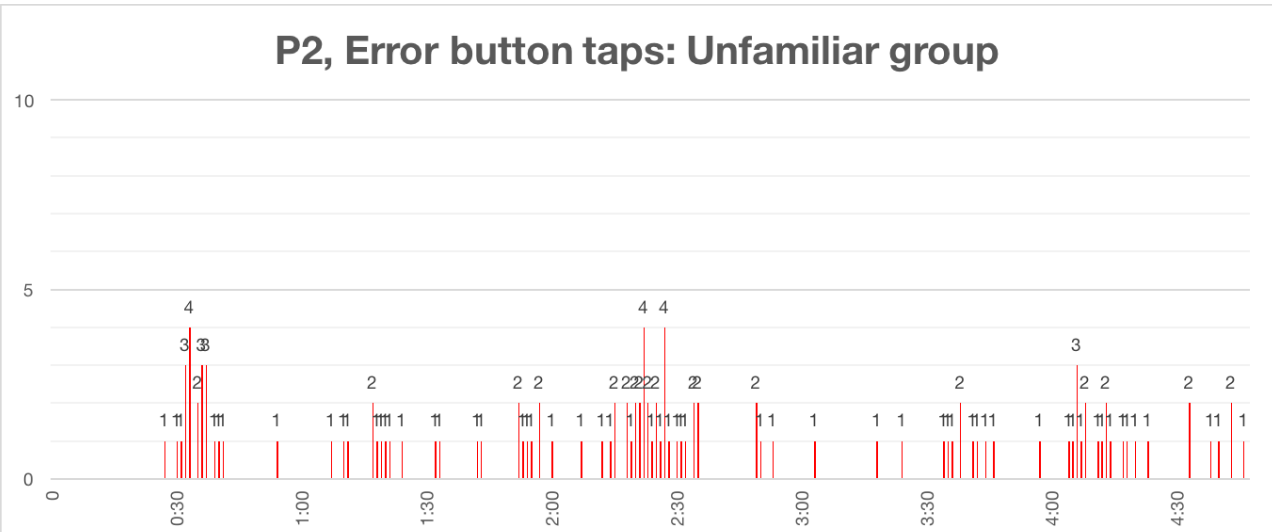


Performance 2: Error
Real-time data visualisation

Tim, experimental

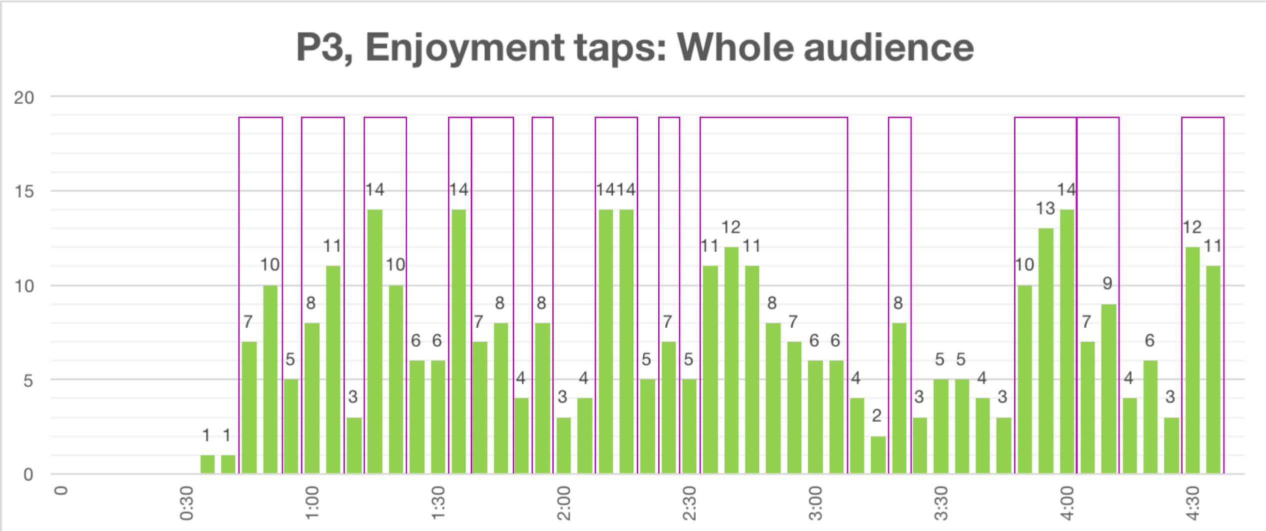


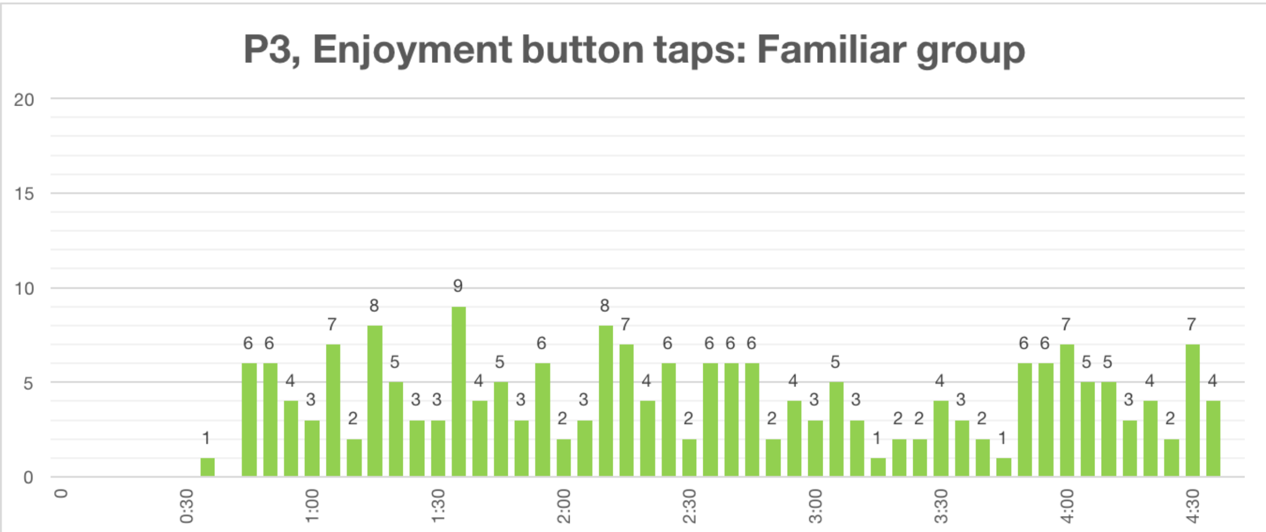


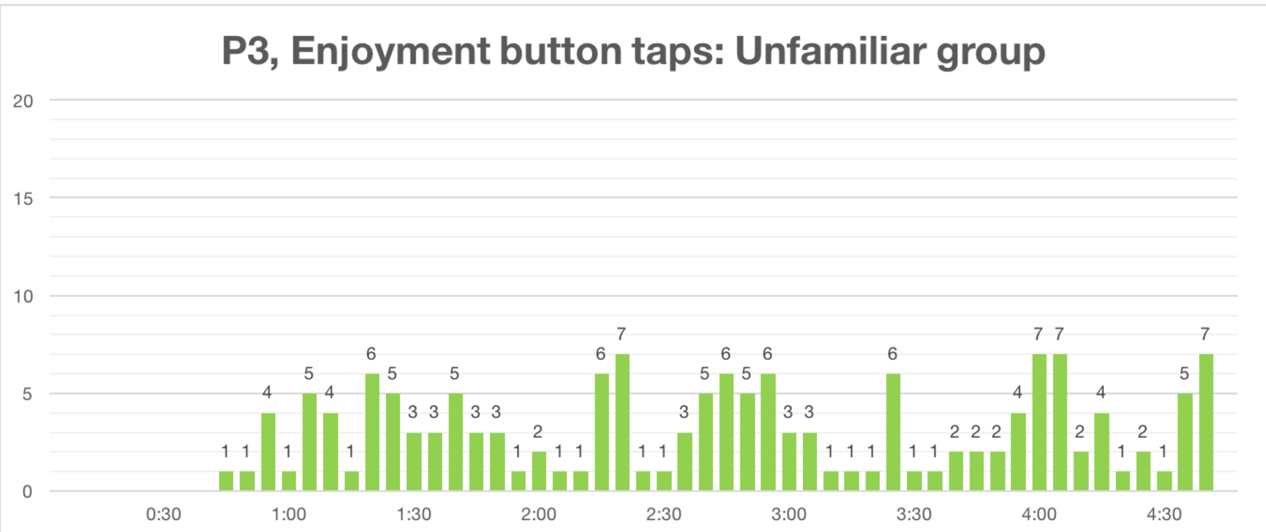


Performance 3: Enjoyment
Real-time data visualisation

Dianne, Conventional

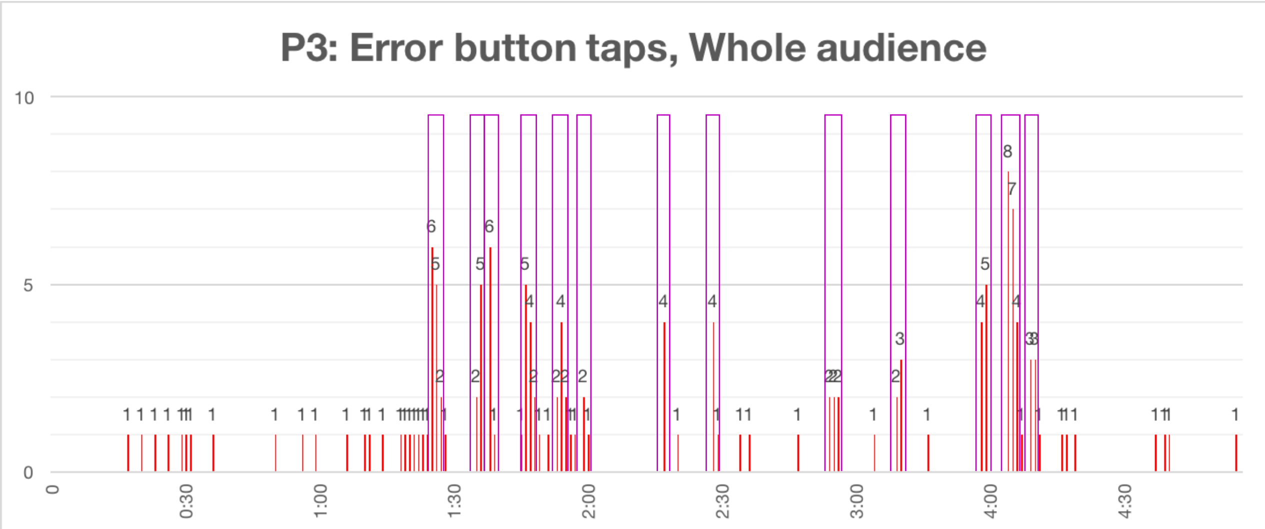


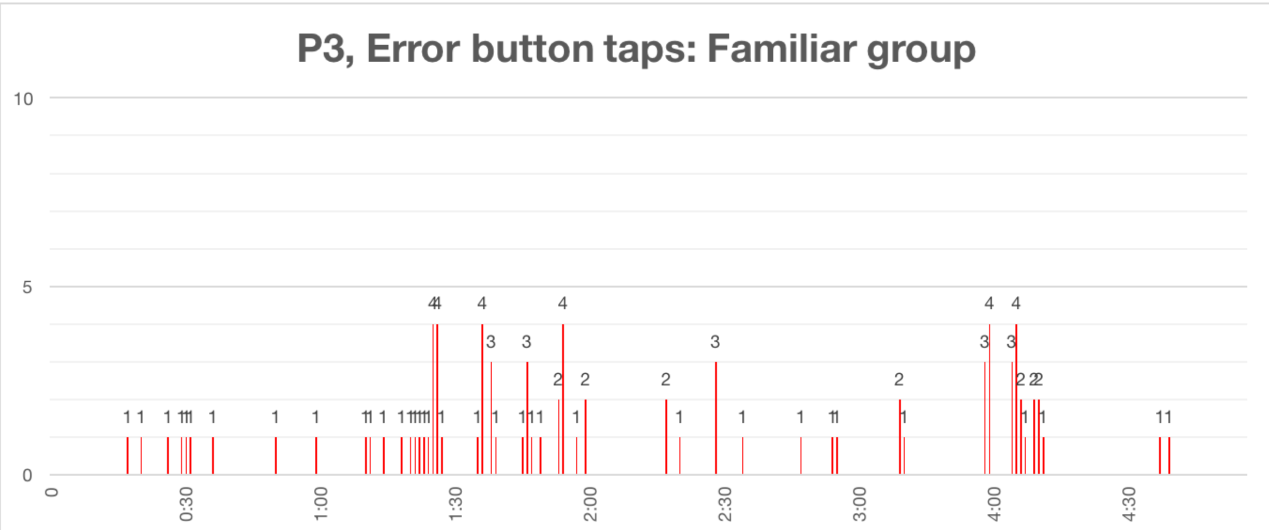


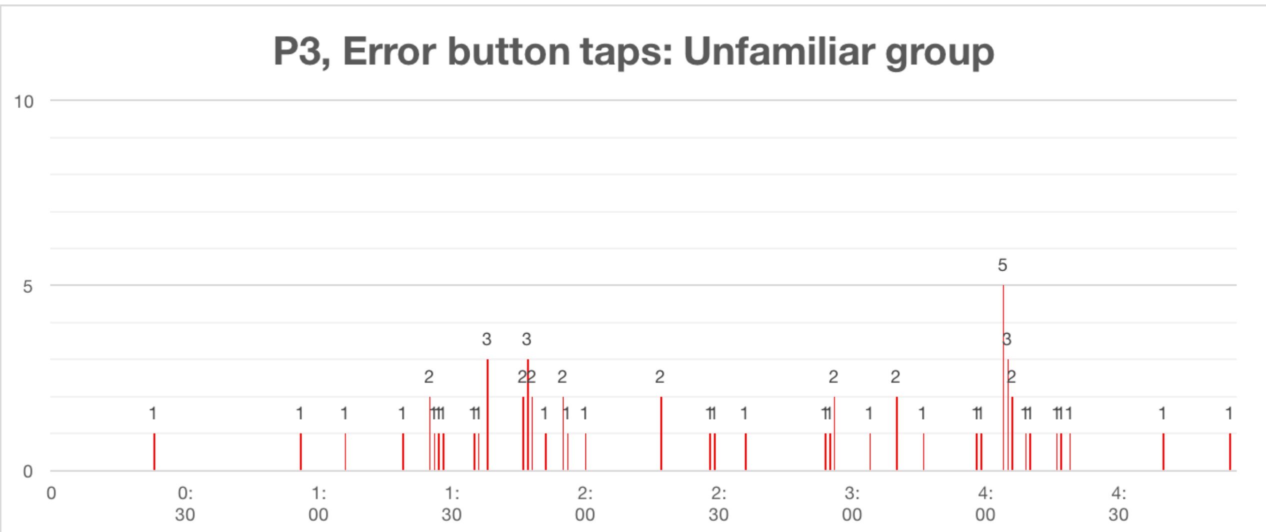


Performance 3: Error
Real-time data visualisation

Dianne, Conventional

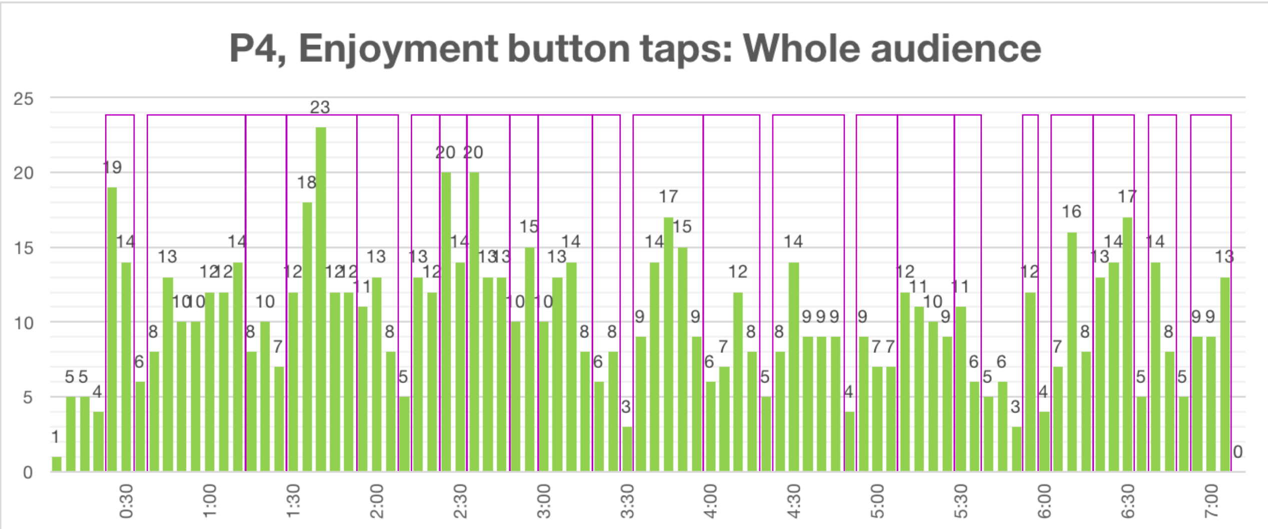


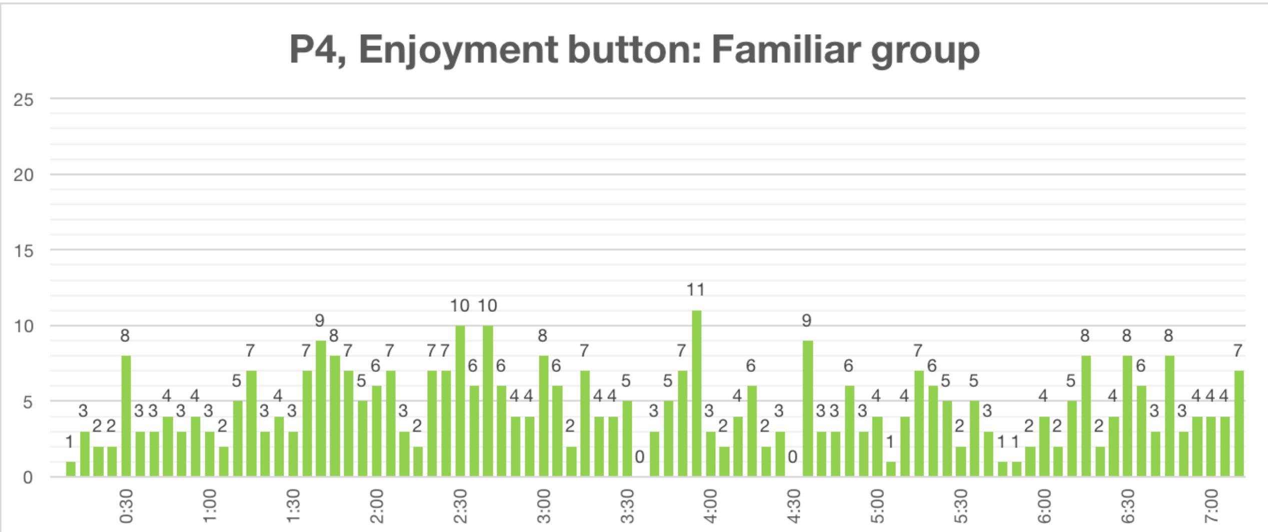


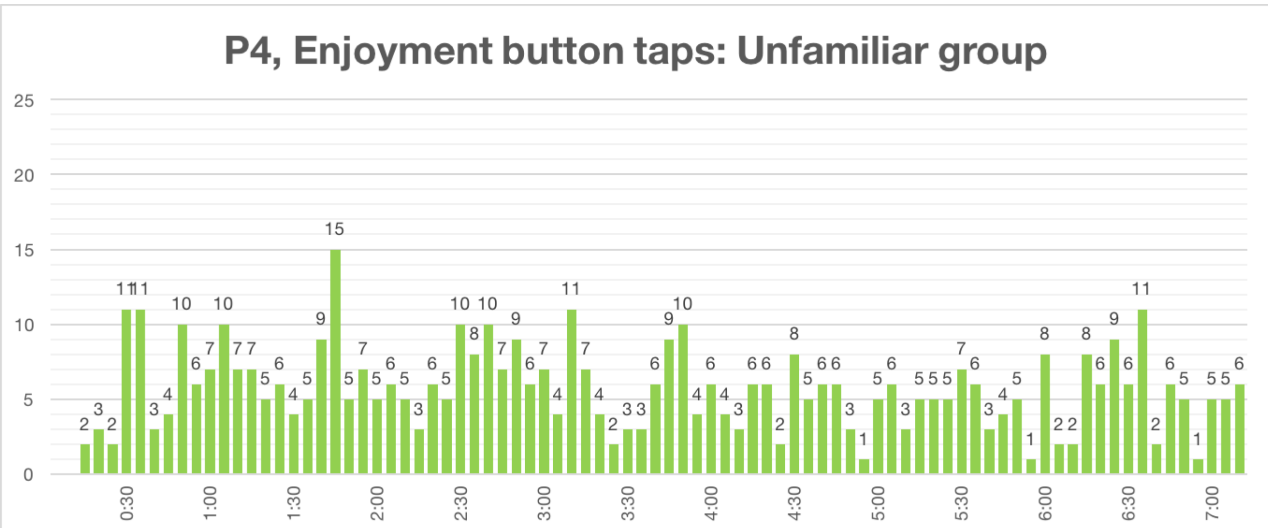


Performance 4: Enjoyment
Real-time data visualisation

Tim, Conventional



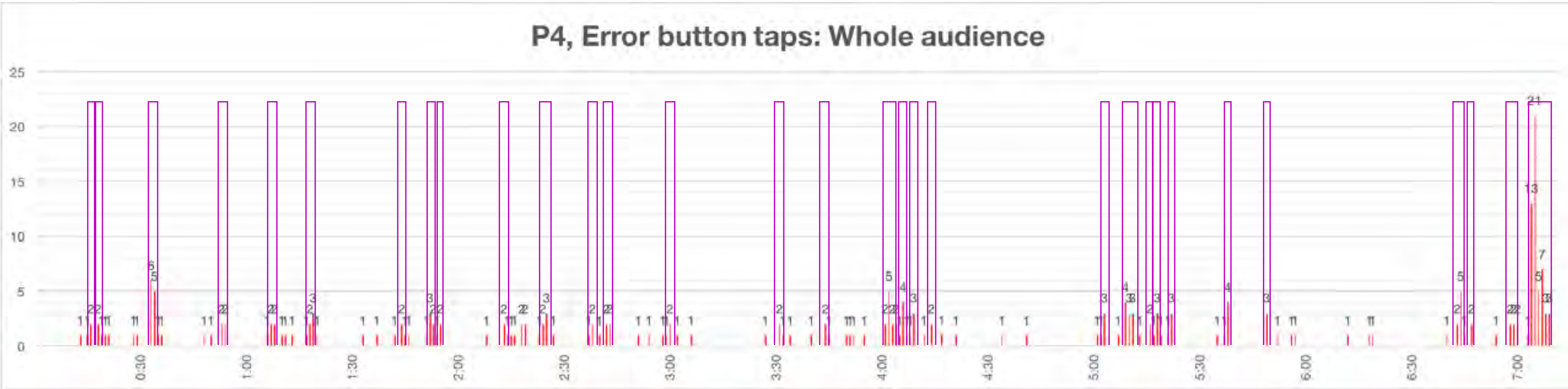


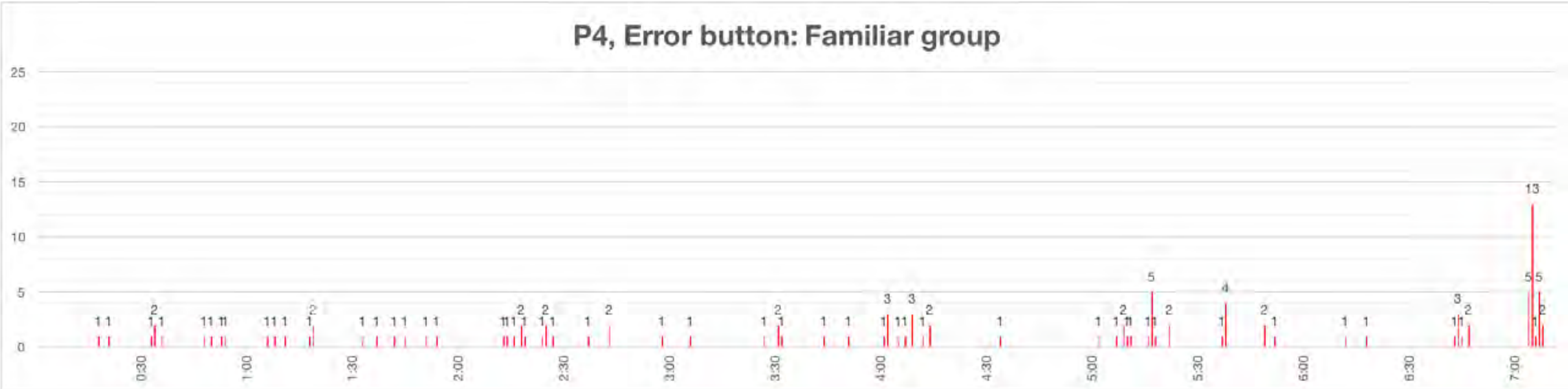


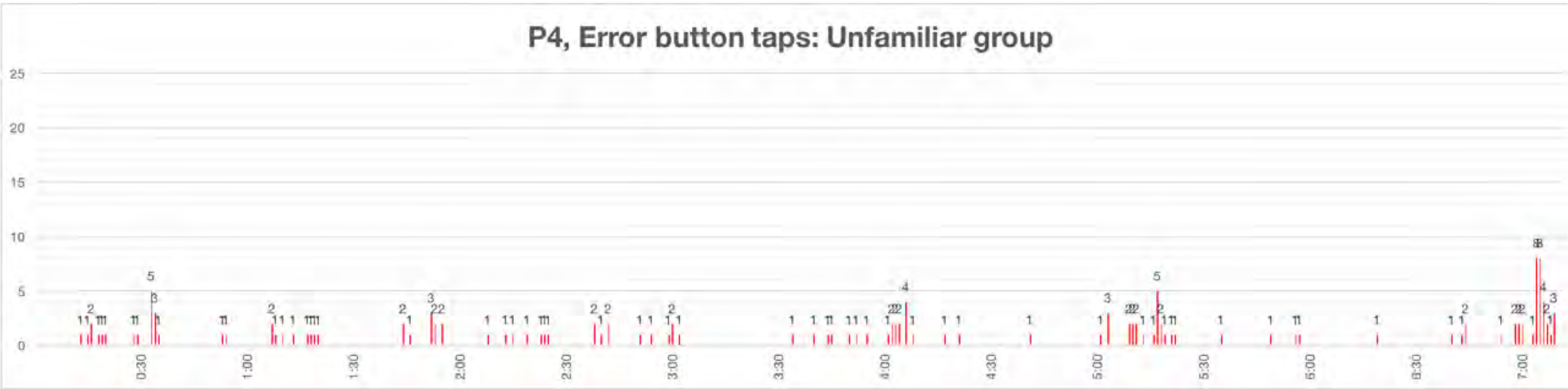
Performance 4: Error

Real-time data visualisation

Tim, Conventional







Appendix F

Study 1: Video documentation

Performance 1: Dianne, Experimental

https://youtu.be/T4CNAo6s92Y

Performance 2: Tim, experimental

https://youtu.be/w10cUzfAU4E

Performance 3: Dianne, conventional

https://youtu.be/d0whdjf87jk

Performance 4: Tim, conventional

https://youtu.be/V_vv-TDdt1s

Performance 1

Dianne, experimental

Enjoyment events						
Time stamp	Avg agreement per bin	Event Num	Features			
0:35-0:50	9.13	1	Three instances of build up/resolution (tapping, then vibrato)	Resolution		
0:50-1:10	12.10	2	End of third instance	End		
1:20-1:25	11.10	3	Changes from playing with stick to mallet, intensity	Change		
1:30-1:55	12.12	4	Introduces fast vibrato. Introduction of boing-boing sounds, then fast/intense creation of sounds	New	Intensity	
1:55-2:10	10.13	5	Changes to stick end of mallet, intensity continues, ends with rapid vibrato	Change	Intensity	
2:25-2:35	11.15	6	Slow vibrato at the end of instrument shaking	Resolution		
2:35-2:50	17.13	7	High pitched, staccato sounds start	New		
2:50-3:15	12.10	8	High pitched sounds give way to vibrato, stick incorporated, mallet, low-pitched sounds begin	Change	New	
3:50-4:00	15.10	9	Singing	New		
4:00-4:10	7.10	10	Singing	New		
4:25-4:30	15.10	11	End of performance	End		

Error events						
Time stamp		Event Num	Features			
1:39	2.10	1	?			
1:45	2.10	2	Unclear, but pitch wobbled a little	Pitch		
1:49	4.10	3	Adjustment of shoulder strap	Technical		
1:56	2.10	4	?			
1:59-2:01	2.17	5	?			
2:36	3.10	6	It's not clear if the vibrato is intentional	Intention		
2:51-2:52	3.15	7	It's not clear if the vibrato is intentional	Intention		
3:16	5.10	8	Hit instrument lightly when adjusting strap	Technical	Perf. action	
3:32	3.10	9	Unsure strikes	Unsure		
4:11-4:13	2.13	10	Hit mic stand	Technical	Perf. action	
4:30	10.10	11	After performance ends; unplanned sound	Technical	Perf. action	

Performance 2

Tim, experimental

Enjoyment events						
Time stamp	Avg agreement per bin	Event Num	Features			
0:45-0:55	7.15	1	Theme reaches development.	Establishment		
1:00-1:05	6.10	2	More theme development	Establishment		
1:15-1:25	8.15	3	New motif developing	New		
1:30-1:40	8.15	4	Change in motif	Change		
1:40-1:50	15.10	5	Adding to changed motif, developing	Developing		
1:50-2:05	7.10	6	New sound profile introduced	New		
2:05-2:10	12.10	7	New motif develops on previous motif	New	Complexity	
2:30-2:40	10.15	8	New sounds in a new section of the piece	New		
2:45-2:55	10.10	9	Increase in intensity	Intensity		
3:00-3:15	7.17	10	Introduction of new sounds, high partials, rhythmic parts	New	Rhythm	
3:15-3:40	8.18	11	Develops a recognisable rhythm	Rhythm		

3:40-3:50	9.15	12	Glissando pitch sounds added	New		
3:55-4:00	6.10	13	Reverb added	New		
4:05-4:10	9.10	14	Rhythm slows, reverb is the focus	New		
4:25-4:35	9.15	15	Rhythm dissolves, new bass sounds appear	New		
Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:32-0:34	3.13	1	Sudden volume	Sudden		
0:35-0:37	4.17	2	Sudden volume	Sudden		
0:39-0:40	4.10	3	?			
1:11	2.10	4	Sudden volume/change in sound profile	Sudden		
1:13	2.10	5	Sudden volume	Sudden		
1:17-1:18	3.10	6	Something gets loud, he turns it down	Reaction		
1:32-1:33	2.15	7	Loud noise cuts out	Drop out		
1:43	2.10	8	Loud, harsh sound	Sudden		
1:52-1:53	2.10	9	Sudden volume	Sudden		
1:57	2.10	10	?			
2:14-2:16	2.17	11	He's twiddling knobs and the sound appears to be uncontrollable	Reaction		
2:17-2:19	2.13	12	He's twiddling knobs and the sound appears to be uncontrollable	Reaction		
2:20-2:21	2.15	13	Appears to get it under control, loud sound	Loud		
2:22-2:23	4.15	14	Appears to get it under control, loud sound	Loud		
2:25-2:27	3.17	15	Sound seems to belong to what's just ended	Loud		
2:34-2:35	2.10	16	He appears to be searching for controls	Reaction		
2:49-2:50	3.15	17	Sound gets loud, he is trying to control it	Loud	Reaction	
3:38	3.10	18	Sudden sound from earlier motif, unlike rhythmic section that's happening	Sudden		
3:41	2.10	19	Sudden sound from earlier motif, unlike rhythmic section that's happening	Incongruency		
4:06-4:08	3.13	20	Harsh sound	Harsh		
4:11	2.10	21	Sound is repeating, appears to be stuck in a process	Control		
4:13	2.10	22	He appears to be searching for controls	Reaction		
4:15	3.10	23	He appears to be searching for controls	Reaction		
4:17-4:18	2.10	24	Twiddles knobs and things get muted but it doesn't seem intentional	Volume		
4:33	3.10	25	Little beep at the end	Sound out of place		
4:38	3.10	26	Turning down knob	Reaction		
4:40	3.10	27	Turning it off	End		

Performance 3

Dianne, conventional

Enjoyment events						
Time stamp	Avg agreement per bin	Event Num	Features			
0:45-0:55	8.15	1	Established rhythm and melody	Establish		
1:00-1:10	9.15	2	Melodic passage, vibrato added	Melodic	Added feature	
1:15-1:25	14.10	3	Melodic passage, started using stick end	Melodic	Change	Flow
1:35-1:40	7.10	4	Melodic passage	Melodic	Flow	
1:40-1:50	7.15	5	Melodic passage, switched to mallet	Melodic	Change	Flow
1:55-2:00	4.10	6	Rhythmic flow	Flow	Rhythm	

2:10-2:20	14.10	7	Rhythmic flow with mallet scrape	Flow	Rhythm	
2:25-2:30	3.15	8	Rhythmic flow	Flow	Rhythm	
2:35-3:10	30.15	9	Intensity of singing	Intensity		
3:20-3:25	4.10	10	Rhythmic flow with vibrato	Flow	Rhythm	
3:50-4:05	18.15	11	Build up of intensity	Intensity		
4:05-4:15	8.10	12	Maintaining of that intensity	Intensity		
4:30-4:40	11.15	13	End of performance	End		

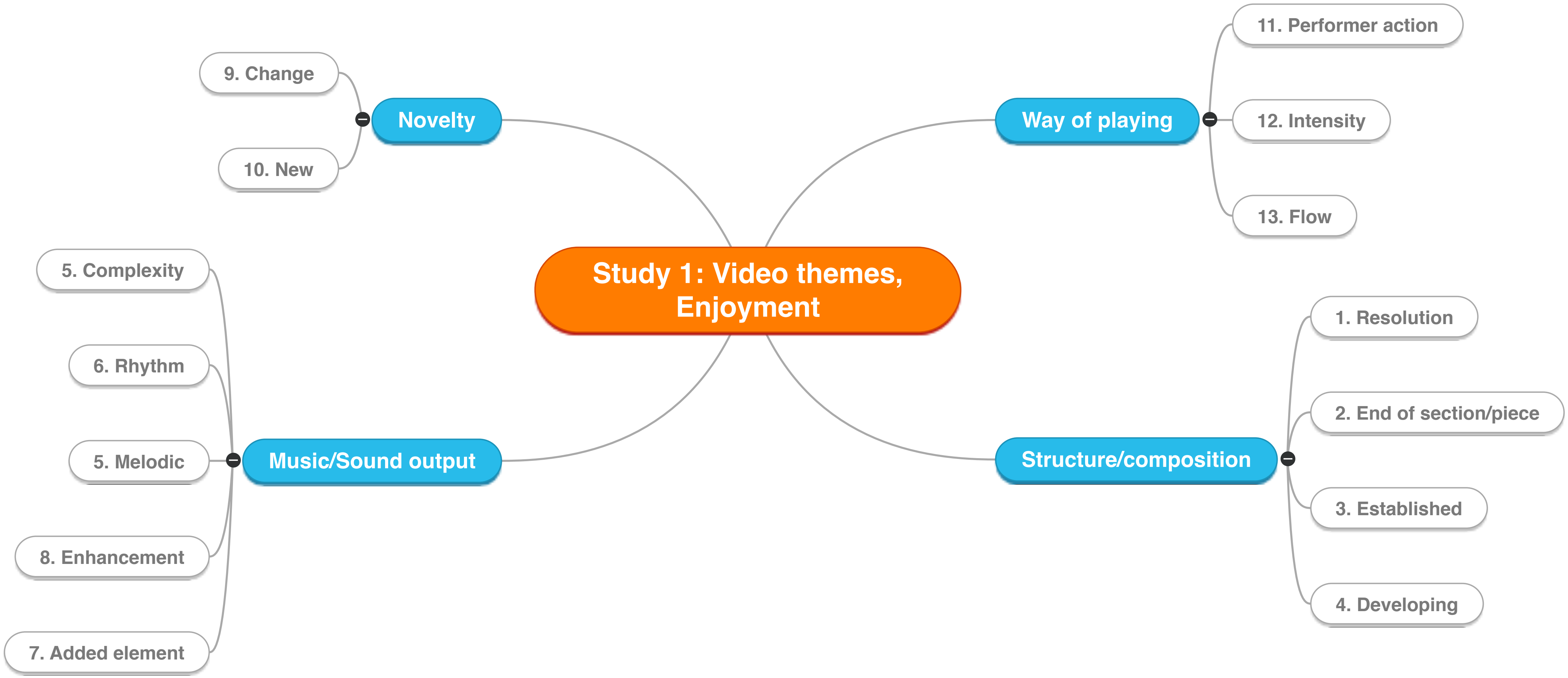
Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
1:25-1:27	4.13	1	Mis-hit with stick	Action		
1:35-1:36	3.15	2	Pitch correction	Pitch		
1:38	6.10	3	Pitch correction	Pitch		
1:46-1:48	3.17	4	?			
1:53-1:55	2.17	5	Hesitant strike	Hesitate		
1:59	2.10	6	Pitch wobble	Pitch		
2:17	4.10	7	Hesitant scrape with mallet	Hesitate		
2:28	4.10	8	Mis-scrape, pitch wobble	Action	Pitch	
2:54-2:56	2.10	9	Hesitant scrape, pitch wobble	Hesitate	Pitch	
3:09-3:10	2.15	10	Low-end pitch correction	Pitch		
3:28-3:29	4.15	11	Mis-hit and pitch correction	Pitch	Action	
3:34-3:36	6.13	12	Pitch correction	Pitch		
3:39-3:40	3.10	13	Pitch correction	Pitch		

Performance 4

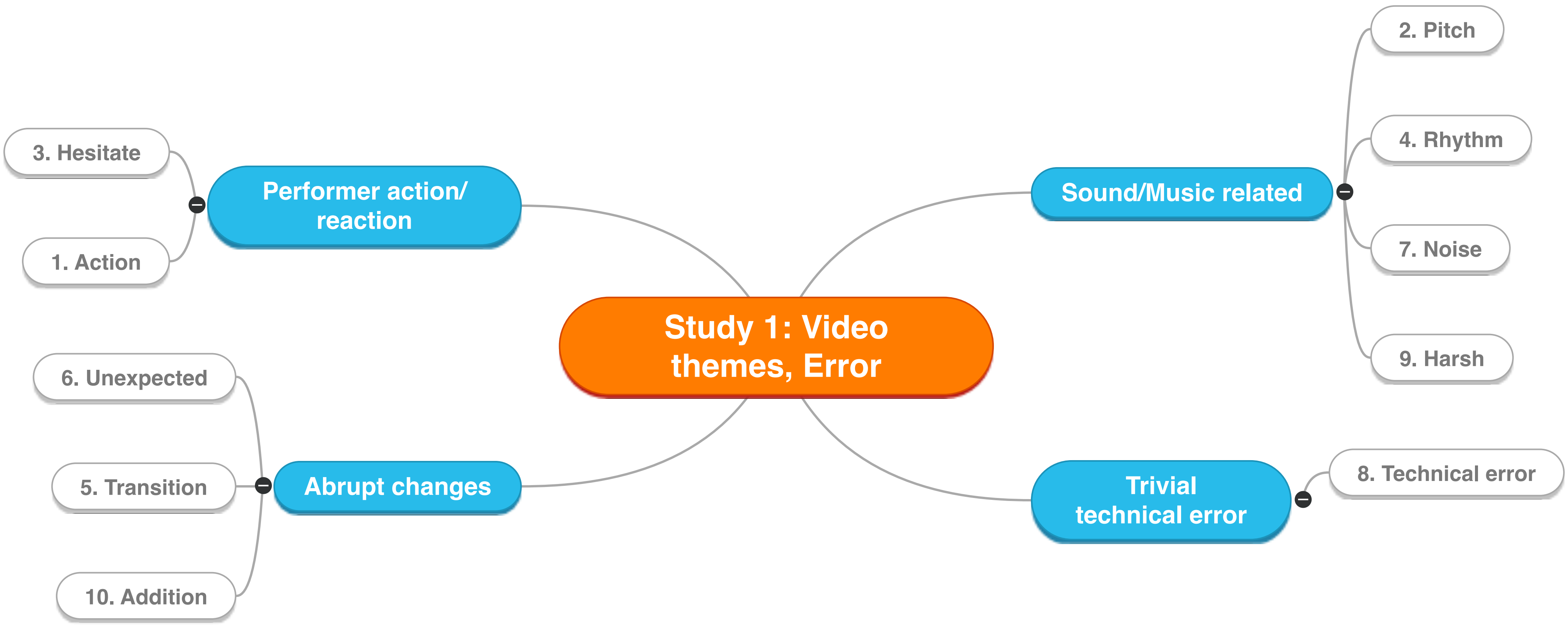
Tim, conventional

Enjoyment events						
Time stamp	Avg agreement per bin	Event Num	Features			
0:30-0:40	16.15	1	Groove established	Establish	Rhythm	
0:45-1:15	9.19	2	Adding vocals, rhtym, game sounds	Addition	Rhythm	
1:15-1:30	8.13	3	Inputting rhythm on keyboard, flow of rhythm	Musical Action	Rhythm	Flow
1:30-1:45	15.14	4	Melodic elements, Tim moving, groove	Perf. action	Flow	Melodic
1:55-2:05	10.17	5	Groove, flow	Flow		
2:15-2:25	12.15	6	Small melodic elements, intense groove	Flow	Addition	
2:25-2:35	17.10	7	High pitched reverb part enters	Addition		
2:35-2:50	15.13	8	Messing with high pitched reverb, groove	Flow	Enhancement	
2:50-3:00	12.15	9	Knob twisting, physically responding	Flow	Perf. action	
3:05-3:15	8.18	10	Moving to beat, flow	Flow	Perf. action	
3:20-3:30	7.10	11	Plays keybord, flow	Flow	Mus action	
3:35-4:00	12.18	12	Adds keyboard rhythm, intensely busy with hands, dancing	Perf. Action	Musical action	
4:00-4:20	8.13	13	Breakup of rhythm, change in motif	Change		
4:25-4:50	9.18	14	Dancing, flow, washy elements come in	Addition	Flow	Perf action
5:10-5:25	7.17	15	Change of motif, hand-entering drum sounds	Change	Mus action	
5:30-5:40	8.15	16	Working sliders with hands, flow	Flow	Mus action	
5:55-6:00	12.10	17	Layering of rhyhtm	Addition		
6:05-6:20	10.13	18	Layering intensifies, different sounds added	Addition		
6:20-6:30	14.17	19	Reverb on new sounds, complexity increases	Complexity	Effects	
6:40-6:50	11.10	20	Into it, dancing, lots of hands, lots of things going on	Complexity	Perf. action	
6:55-7:10	10.13	21	Intensity, switches off	Intensity	End	

Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:15	2.10	1	Rhythm inconsistencies	Rhythm		
0:17	4.10	2	Rhythm inconsistencies	Rhythm		
0:33-0:34	5.15	3	Rhythm inconsistencies	Rhythm		
0:53-0:54	2.10	4	?			
1:07-1:08	2.10	5	Adding of rolling rythm	Unexpected	New	
1:18-1:19	2.15	6	Some input rhythm was inconsistent	Perf. action	Rhythm	
1:44	2.10	7	Hand-inputting rhth	Perf. action	Rhythm	
1:52-1:53	2.15	8	Transition inconsistency	Transition		
1:55	2.10	9	Transition inconsistency	Transition		
2:13	2.10	10	Unexpected sound	Unexpected	Sound	
2:18-2:19	2.10	11	?			
2:24-2:25	2.15	12	Unexpected inconsistency	Unexpected	Inconsistency	
2:38	2.10	13	Unexpected inconsistency	Unexpected	Inconsistency	
2:42-2:43	2.10	14	Unexpected inconsistency	Unexpected	Inconsistency	
3:00	2.10	15	Unexpected inconsistency	Unexpected	Inconsistency	
3:31	2.10	16	?			
3:44	2.10	17	?			
4:01-4:05	2.18	18	Stops rhythm with harsh noise	Stop	Harsh	
4:06	4.10	19	Harsh noise again	Harsh		
4:09	3.10	20	Harsh noise again	Harsh		
4:14	2.10	21	Harsh noise again	Harsh		
5:03	3.10	22	Rhythm inconsistencies	Inconsistency	Rhythm	
5:09-5:11	3.13	23	Rhythm inconsistencies	Inconsistency	Rhythm	
5:17	2.10	24	Tapping in rhythm with hand, not entirely accurate	Inconsistency	Rhythm	Perf. Action
5:19	3.10	25	Tapping in rhythm with hand, not entirely accurate	Inconsistency	Rhythm	Perf. Action
5:23	3.10	26	Adjusting volume with knob, noticeable	Perf. Action		
5:38	4.10	27	Inconsistency may seem like a drop out	Inconsistency		
5:49	3.10	28	Sound starts, he seems surprised	Perf. Action		
6:43-6:44	3.15	29	Rhythm inconsistencies	Inconsistency	Rhythm	
6:47	2.10	30	Breaks up rhythm	Rhythm		
6:58-7:00	2.10	31	Noise introduced into rhythm	Noise	Change	
7:04-7:09	8.15	32	Turn off	Technical error	Perf. Action	

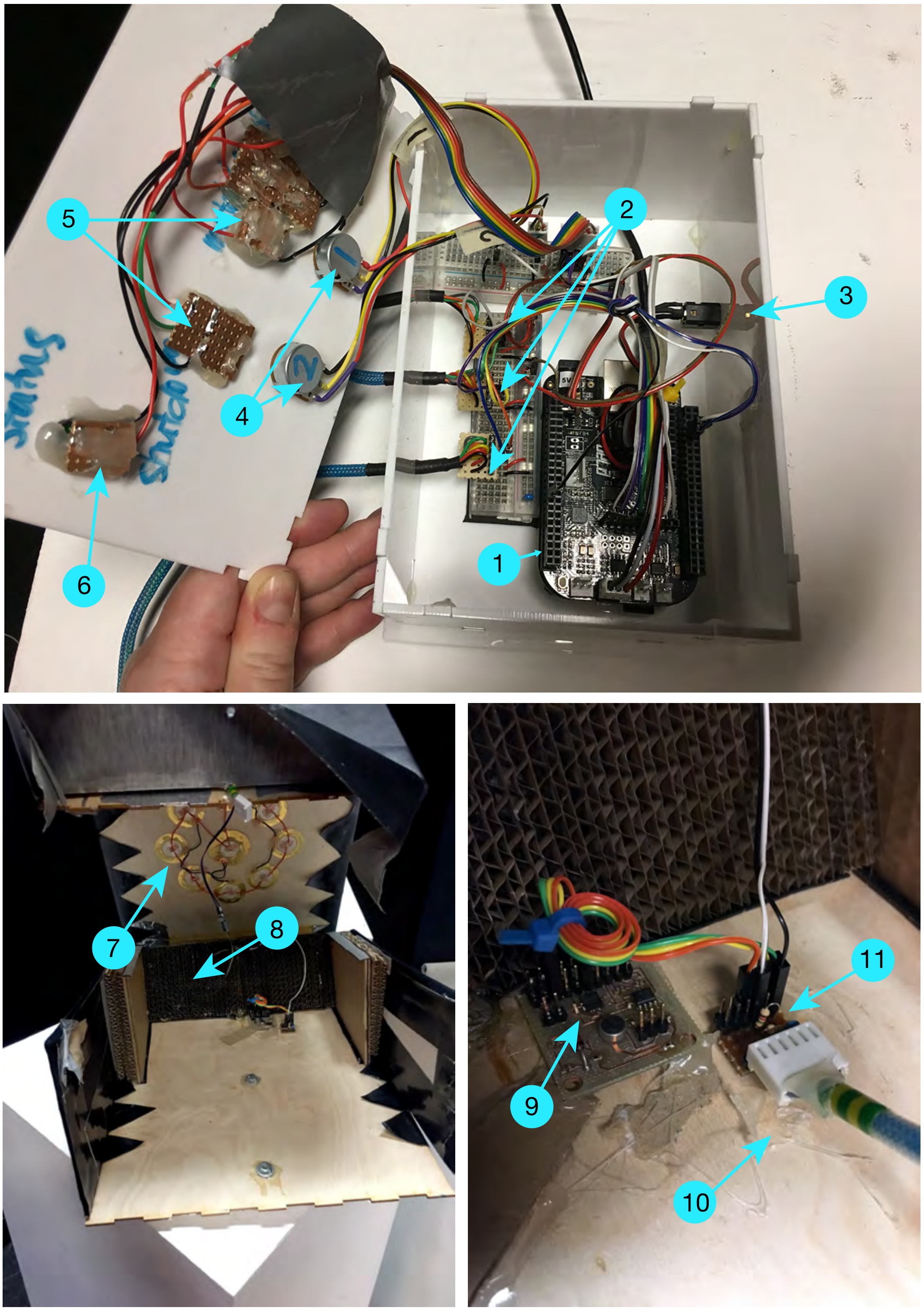


s01_videoThemes_enj l disastrid l 2017-12-06


Study 1: Video themes, Error
Performer action/ reaction
3. Hesitate
1. Action
Abrupt changes
6. Unexpected
5. Transition
10. Addition
Sound/Music related
2. Pitch
4. Rhythm
7. Noise
9. Harsh
Trivial technical error
8. Technical error

Appendix G

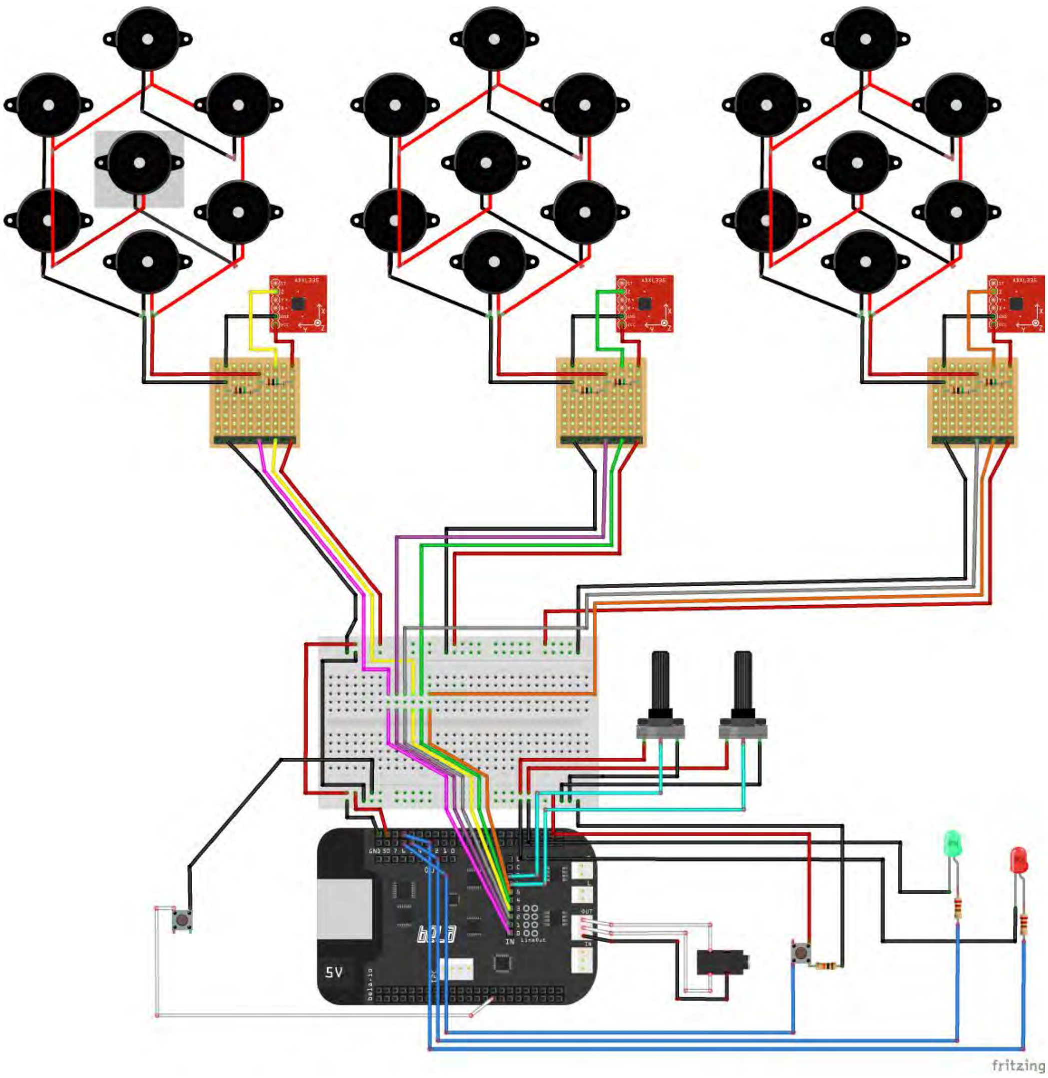
MOAI: Technical diagram



LEGEND

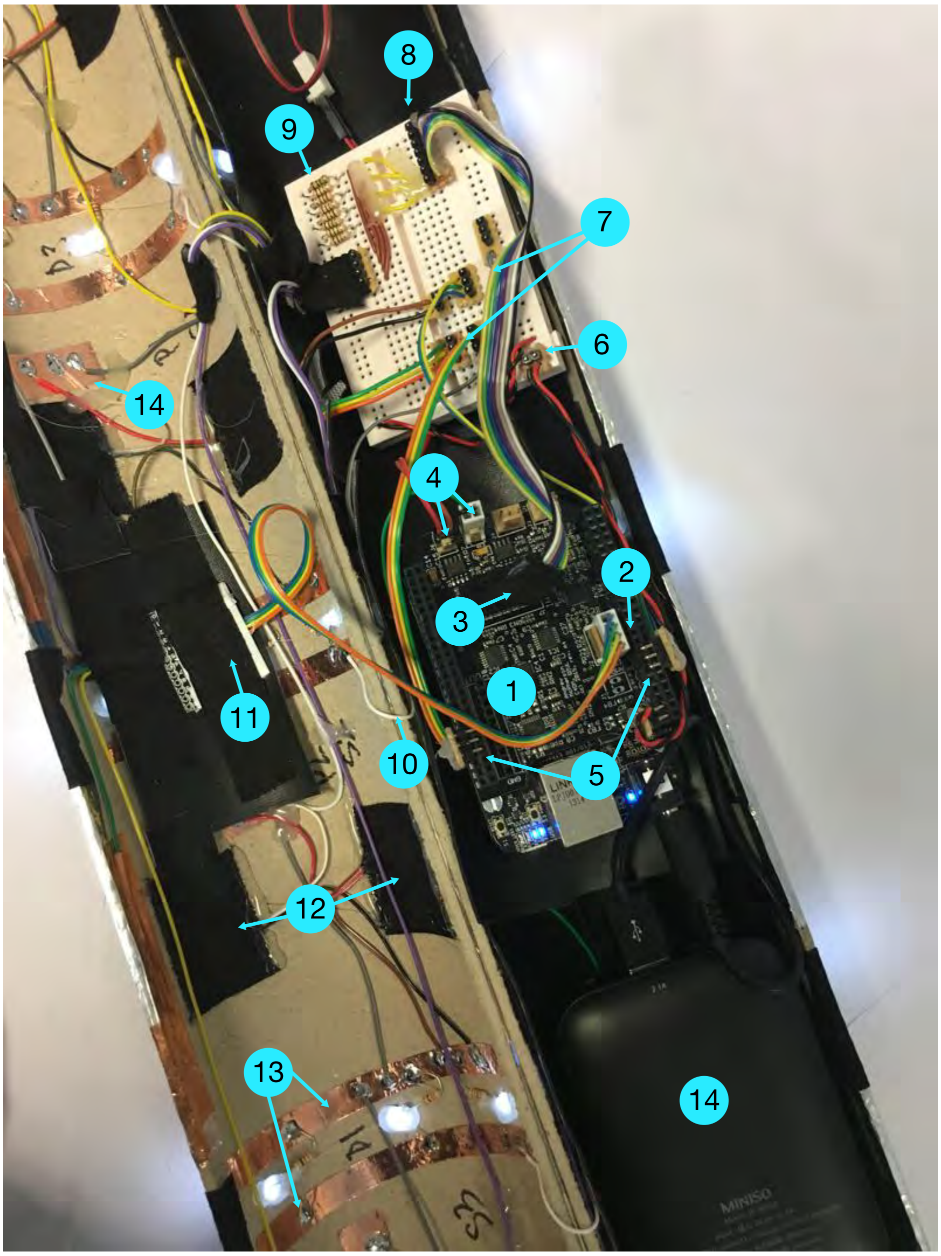
Inside control unit:
1. Bela (embedded audio and sensor proceszsing computer)
2. 3 cables that connect element outputs to Bela inputs (piezos, accelerometer)
3. Audio out
4. Potentiometers (volume, and strike threshold for sound with high partials)
5. Mute and shutdown buttons
6. Status light (mute light on the other side behind tape)

Inside each element:
7. Piezo network
8. Cardboard core that supports rounded front
9. Accelerometer (ADXL335)
10. Cable that connects sensor outputs to control box
11. Resistor ladder (1MΩ) for piezo signal before being sent to control box


5V
fritzing

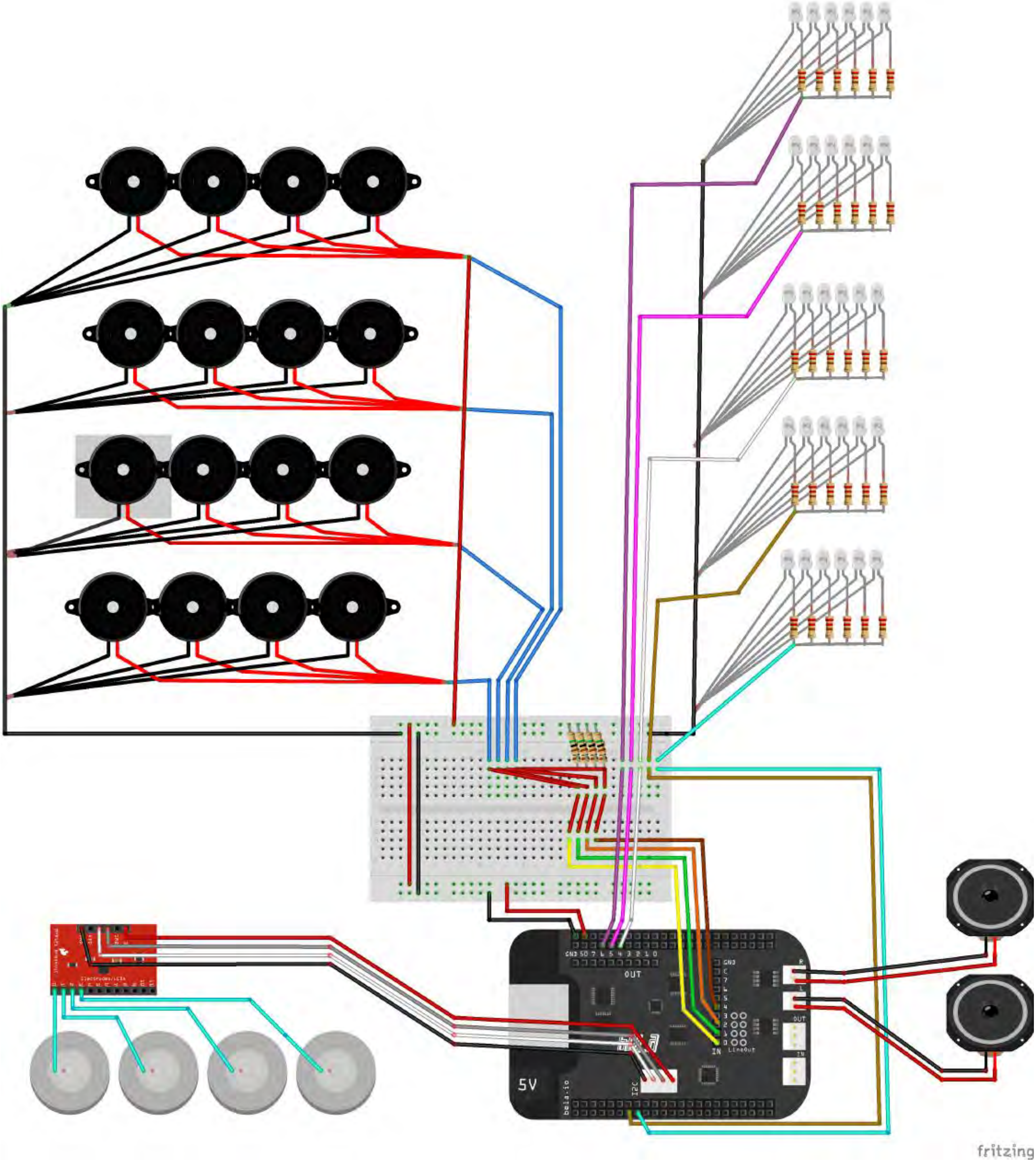
Appendix H

Keppi: Technical diagram



LEGEND:

1. Bela unit
2. i2c connector
3. Analogue input connector
4. Speaker outputs
5. Digital connectors
6. Power connection (3.3V from Bela)
7. Digital connections (from LED collector wires to Bela)
8. Analogue connector to carry 4 piezo inputs to Bela
9. Resistor ladder for piezo inputs (1MΩ)
10. Wire to jump hinge and connect opposite sides of LED group
11. MPR121 board; all electrodes connected to this
12. 2 piezos of 4 piezo network (other 2 on other side of tube)
13. LED group. Upper: Ground. Lower: power. 1 x 220Ω resistor used for each LED.
14. Example of copper tape terminal for half a piezo network.
15. External battery to power speakers.



OUT
5V
bela.io
I2C
IN
fritzing

Appendix I

Study 2: Audience questionnaire

PERFORMANCE 1 (percussion instrument)

1. How much did you enjoy the performance?

1	2	3	4	5
Not at all				Very much

1. How interesting did you find the performance?

1	2	3	4	5
Not at all				Very much

2. Did you understand how the instrument worked?

1	2	3	4	5
Not at all				Very much

3. What did you like about the performance?

4. What did you dislike about the performance?

5. How would you describe the performance to a friend?

PERFORMANCE 2 (laptop)

1. How much did you enjoy the performance?

1	2	3	4	5
Not at all				Very much

2. How interesting did you find the performance?

1	2	3	4	5
Not at all				Very much

3. Did you understand how the instrument worked?

1	2	3	4	5
Not at all				Very much

4. What did you like about the performance?

____

____

____

____

5. What did you dislike about the performance?

____

____

____

____

6. How would you describe the performance to a friend?

____

____

____

____

END OF CONCERT SURVEY

1. What is your age? Please circle one:

Under 18 | 18-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65+ | Prefer not to say

2. What is your gender? Please circle one:

Female | Male | Other | Prefer not to say

3. Are you involved in music in any of the following capacities (tick all that apply, and indicate your number of years of experience):

- ☐ Musician, _________ years
- ☐ Performer, _______ years
- ☐ Composer, _______ years
- ☐ Researcher, _______ years
- ☐ Teacher, ______ years
- ☐ Hobbyist, ______ years
- ☐ Student, ______ years
- ☐ Instrument design, ______ years
- ☐ Other (specify): ____, ______ years

4. Which musical genres do you listen to most frequently? Name up to 3:

____

Thinking about both performances:

Please rate the performances, using **1 (favourite)** and **2 (least favourite):**

___________ **Performance 1 (instrument)** ___________ **Performance 2 (laptop)**

Why did you prefer one performance over the other?

____

____

____

Thinking about Performance 1 (percussion instrument):

What I enjoyed most was:

___

___

___

What I enjoyed least was:

___

___

___

Did you notice any errors? If so, describe them:

___

___

___

Describe, to your knowledge, how the instrument works.

___

___

___

This performance held my attention.

1	2	3	4	5
Not at all				Very much

I understood how the instrument worked.

1	2	3	4	5
Not at all				Very much

I could play this instrument.

1	2	3	4	5
Not at all				Very much

Thinking about Performance 2 (laptop):

What I enjoyed most was:

___

___

___

What I enjoyed least was:

____

____

____

Did you notice any errors? If so, describe them:

____

____

____

Describe, to your knowledge, how the instrument works.

____

____

____

This performance held my attention.

1	2	3	4	5
Not at all				Very much

I understood how the instrument worked.

1	2	3	4	5
Not at all				Very much

I could play this instrument.

1	2	3	4	5
Not at all				Very much

Finally ....

How would you describe your experience using the mobile phone interface? Do you have any suggestions to make it better?

____

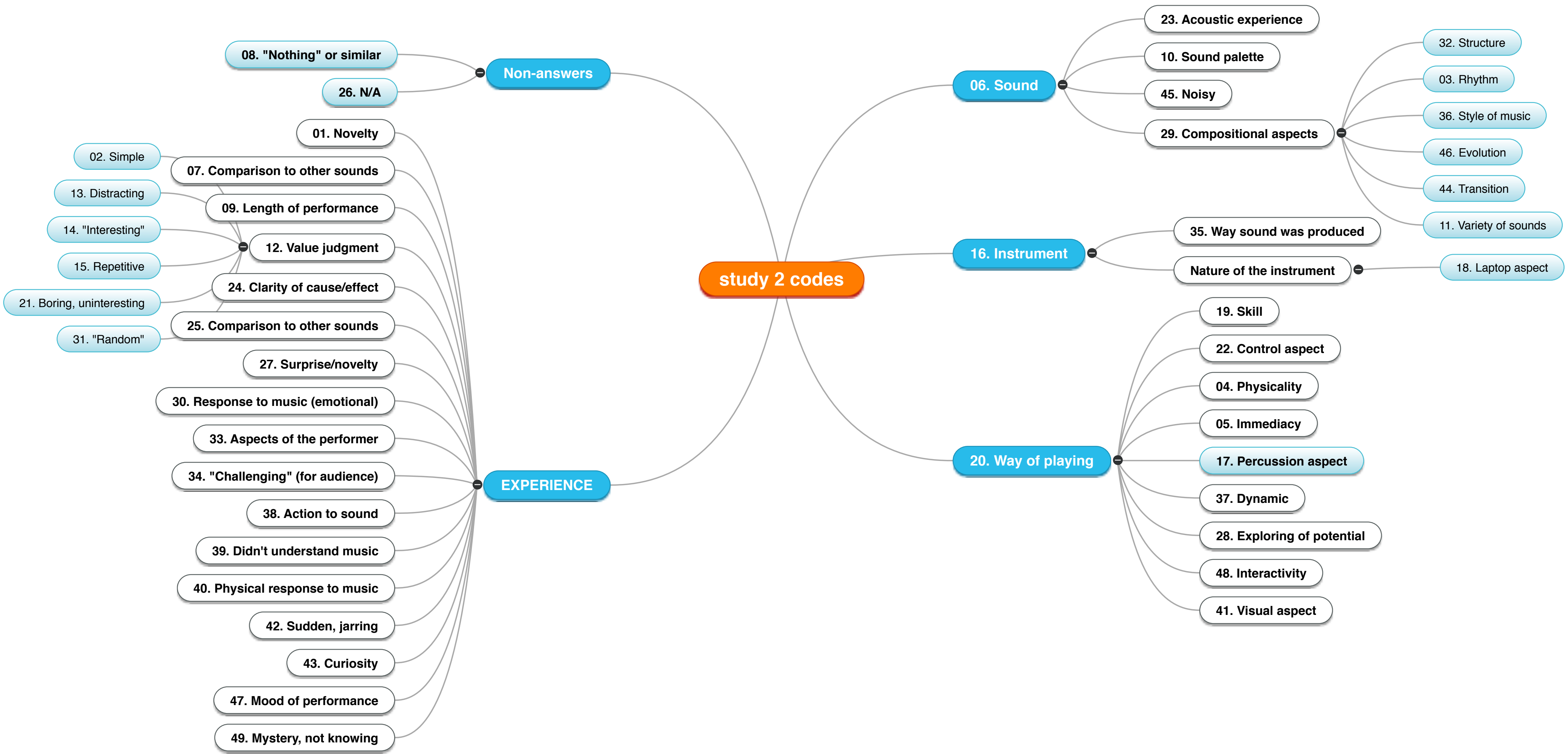
____

____

Thank you so much for your participation!

Appendix J

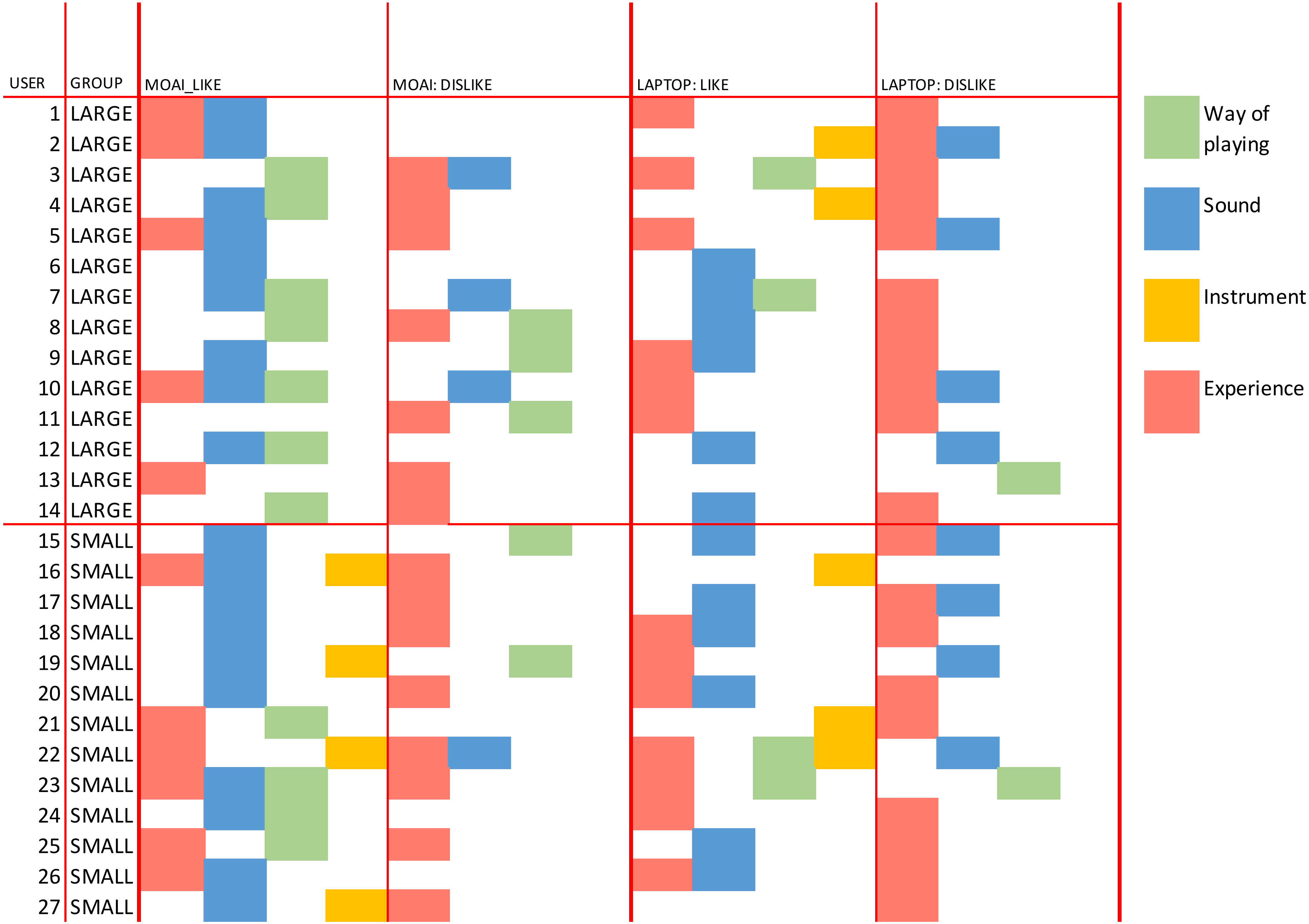
Study 2: Qualitative analysis



code_map_s02 I disastrid I 2017-11-02

USER	GROUP	MOAI_LIKE	MOAI_DISLIKE	LAPTOP_LIKE	LAPTOP_DISLIKE
1	LARGE	1, 3		34	34, 43
2	LARGE	3, 9		35	12, 36, 41
3	LARGE	5	11, 12	14, 37	38
4	LARGE	3, 17	13	35	39
5	LARGE	14, 6	15	40	36, 41
6	LARGE	6	8	10	26
7	LARGE	6, 17	23	22, 36	41, 43
8	LARGE	19, 20	21, 20, 22	3	41
9	LARGE	16, 6	20	27, 6, 30	42
10	LARGE	4, 3, 24	11 A	27	21, 29
11	LARGE	26	25, 20	43	21
12	LARGE	6, 4	26	6	32
13	LARGE	27	15	26	4 A
14	LARGE	20	15	6	41, 43
15	SMALL	6, 3	28	3	41, 44
16	SMALL	16, 29, 30	15	35	
17	SMALL	3	31	3, 46	45, 41
18	SMALL	3	21	3, 14, 33	21
19	SMALL	16, 29	28	27, 2	11
20	SMALL	32	15	30, 3	9
21	SMALL	2, 4		35, 16	47
22	SMALL	16, 14	29, 21	14, 35, 22	21, 6
23	SMALL	33, 20, 6	21	22, 14	48 A
24	SMALL	20, 3, 6	8	49	8
25	SMALL	12, 20	15	3	41
26	SMALL	14, 6, 24		3, 14	41
27	SMALL	6, 16	15		39

EXPERIENCE	1, 7, 9, 24, 25, 27, 30, 33, 34, 38, 39, 40, 42, 43, 47, 49
EXPERIENCE: Value judgments	12, 2, 13, 14, 15, 21, 31, 41
INSTRUMENT	16, 18, 35
SOUND	6
SOUND: Compositional	29, 3, 11, 32, 36, 44, 46
SOUND: Qualities	10, 23, 45
Way of playing	20, 4, 5, 17, 19, 22, 28, 37, 48
Non-answers	8, 26

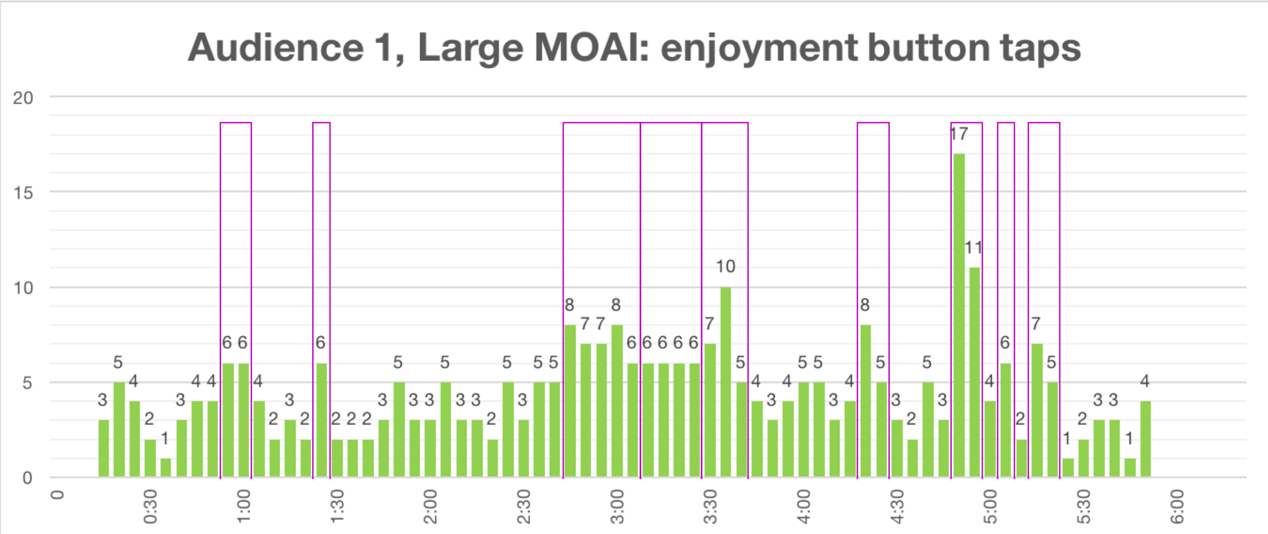


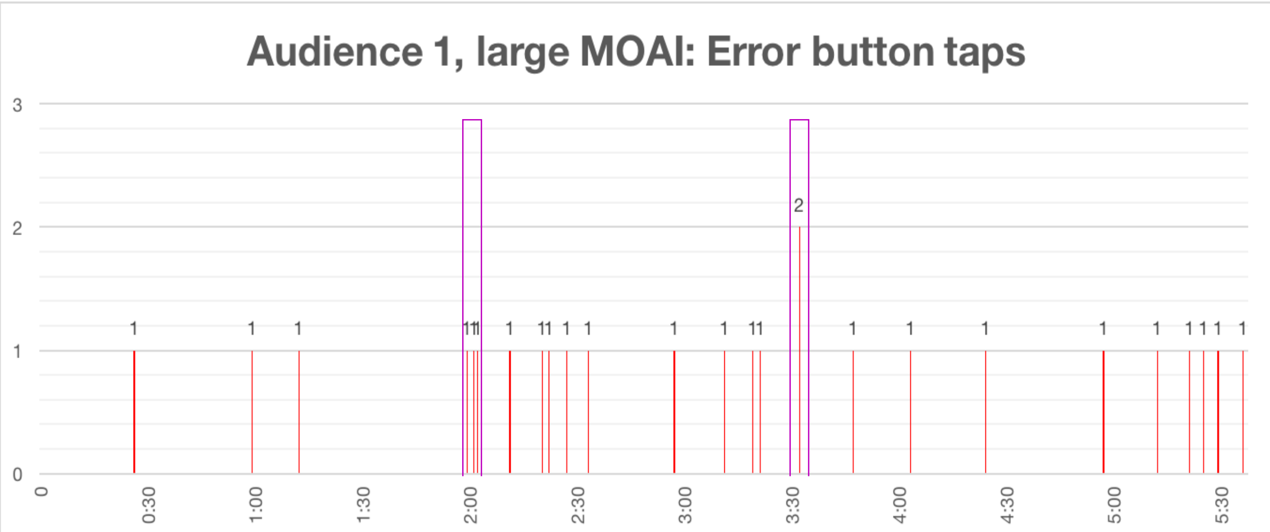
USER
GROUP
MOAI_LIKE
MOAI: DISLIKE
LAPTOP: LIKE
LAPTOP: DISLIKE
1 LARGE
2 LARGE
3 LARGE
4 LARGE
5 LARGE
6 LARGE
7 LARGE
8 LARGE
9 LARGE
10 LARGE
11 LARGE
12 LARGE
13 LARGE
14 LARGE
15 SMALL
16 SMALL
17 SMALL
18 SMALL
19 SMALL
20 SMALL
21 SMALL
22 SMALL
23 SMALL
24 SMALL
25 SMALL
26 SMALL
27 SMALL
Way of playing
Sound
Instrument
Experience

Appendix K

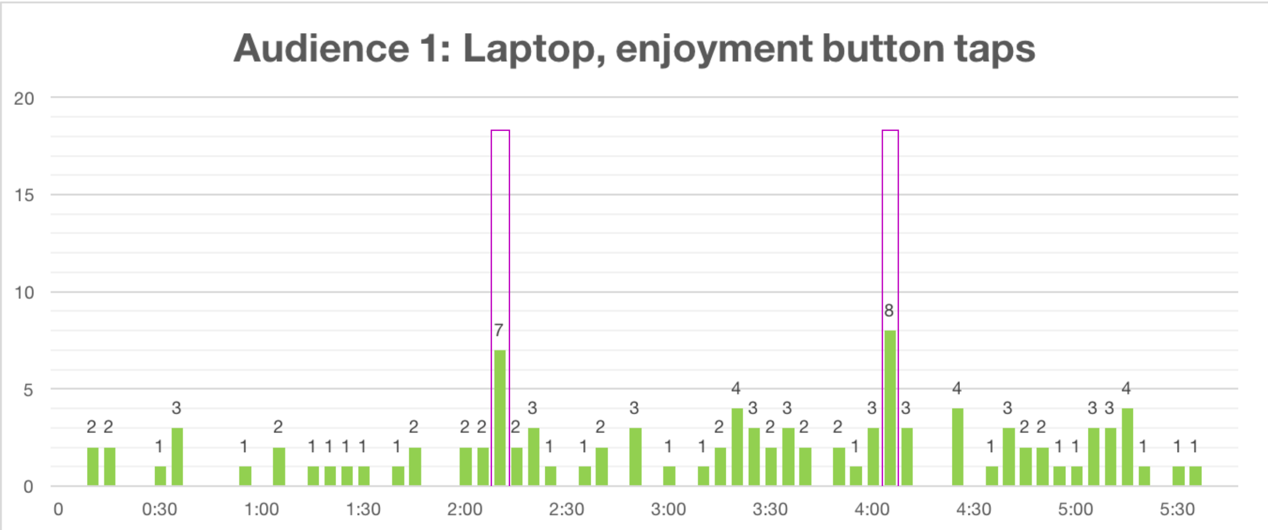
Study 2: Real-time analysis

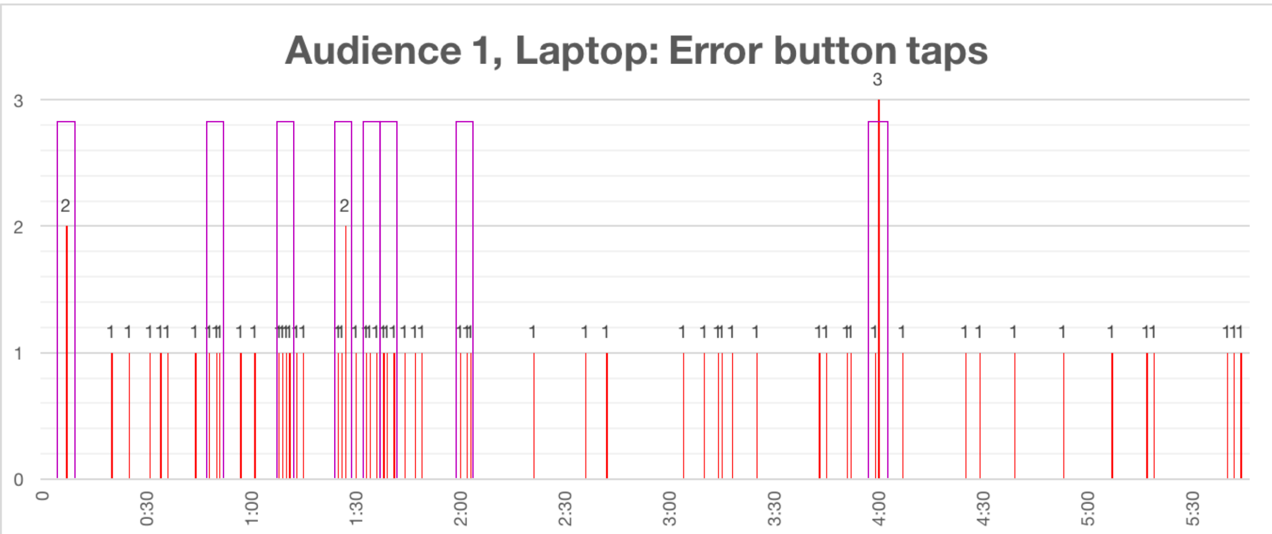
Audience 1: Large MOAI
Real-time data visualisation



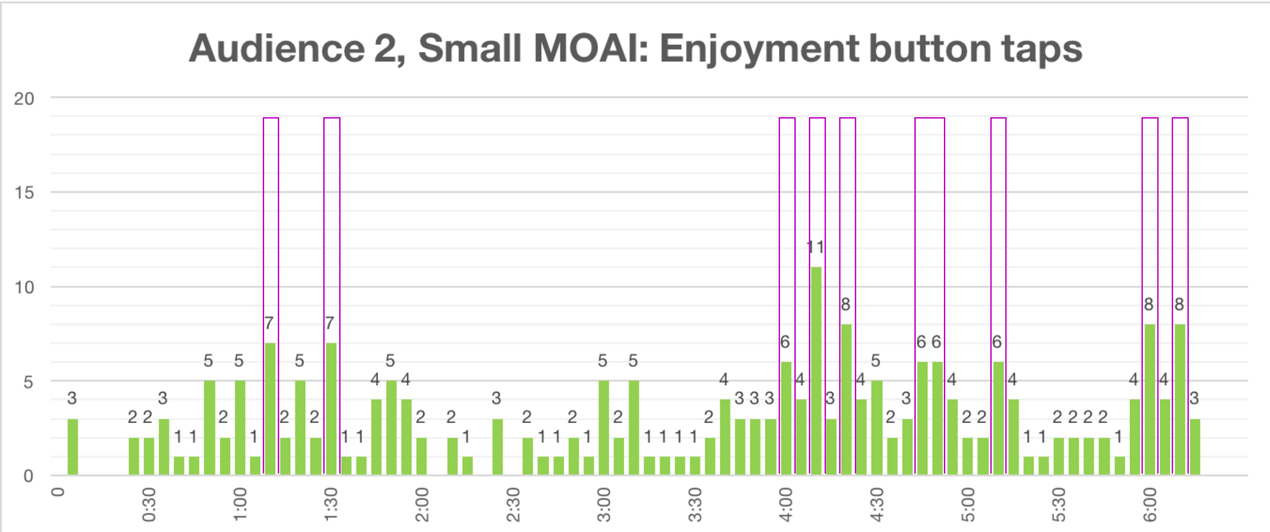


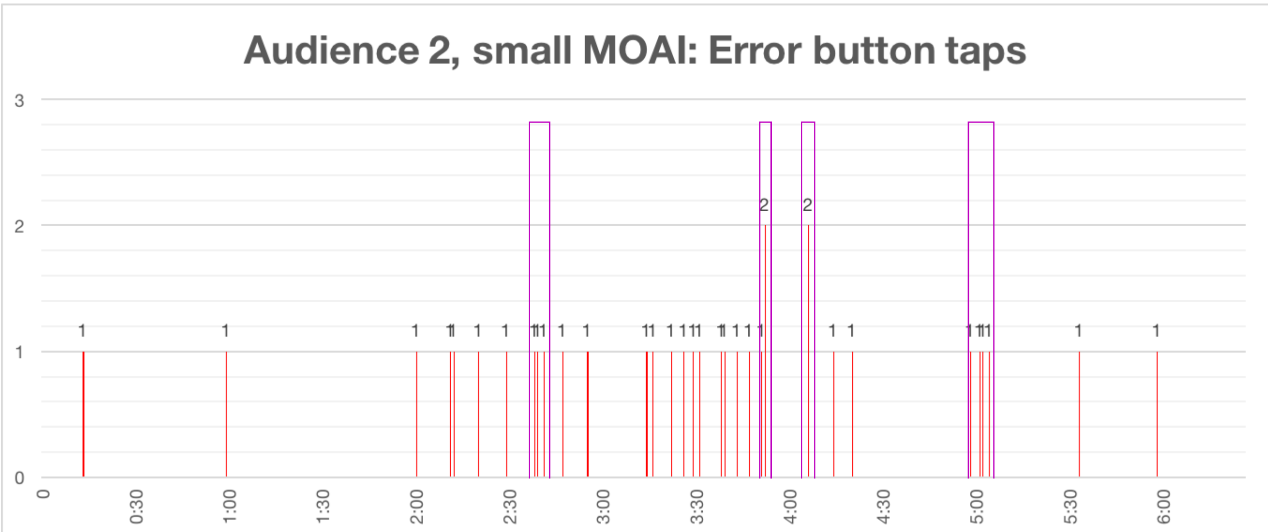
Audience 1: Laptop
Real-time data visualisation



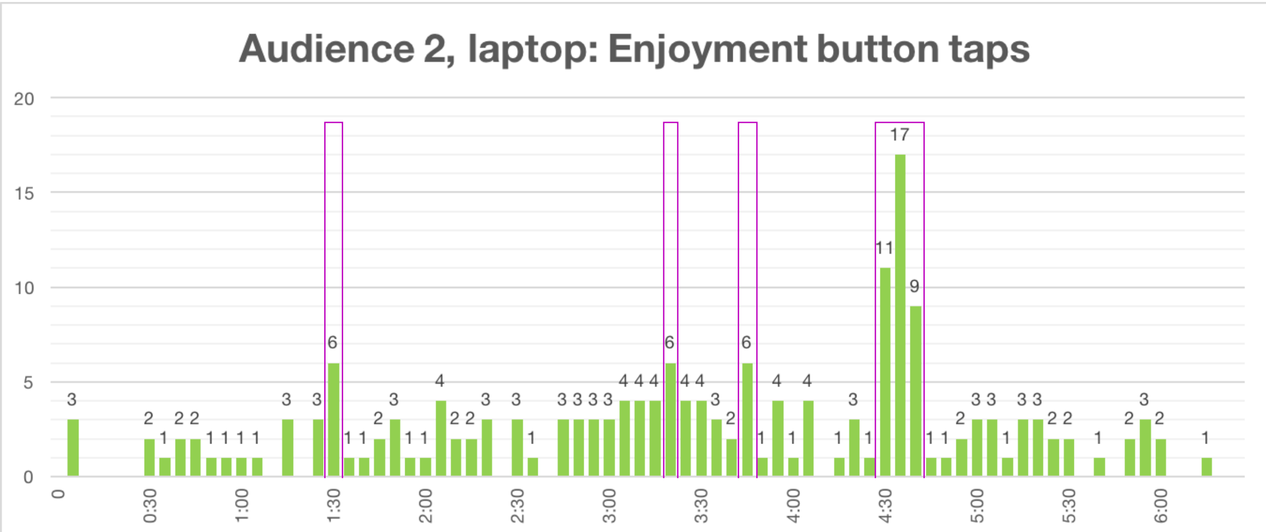


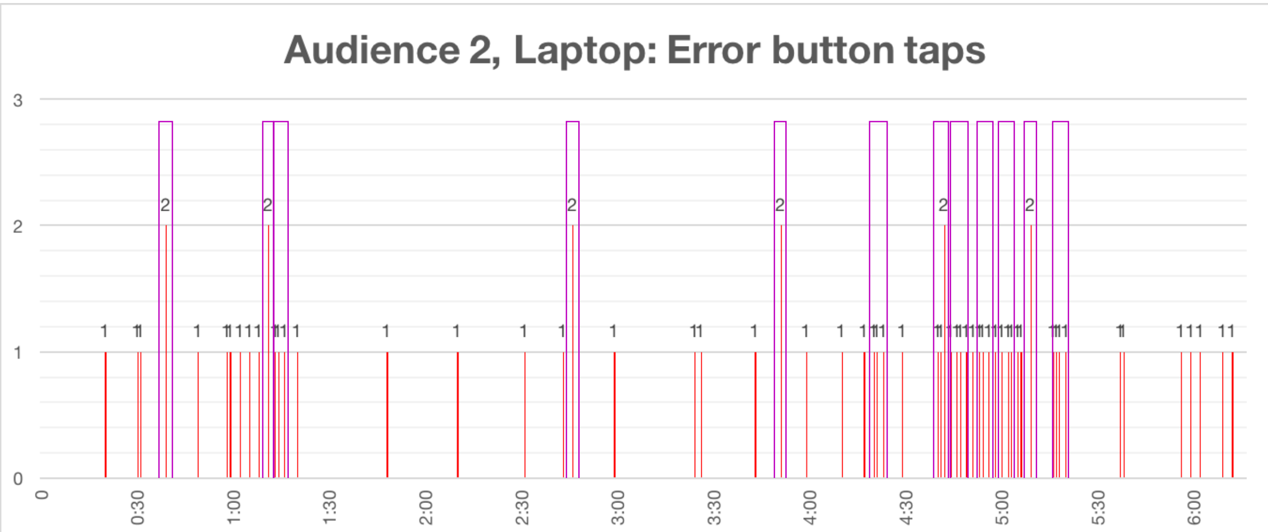
Audience 2: Small MOAI
Real-time data visualisation





Audience 2: Laptop
Real-time data visualisation





Appendix L

Study 2: Video documentation

Audience 1

MOAI (Large)

https://youtu.be/TWIrZ2AYa_w

Laptop Performance

https://youtu.be/Mc2_b9QWXM4

Audience 2

MOAI (Small)

https://youtu.be/puaU3HTdRHg

Laptop Performance

https://youtu.be/uN4Q5xwa7NI

Audience 1

MOAI (Large)

Enjoyment events

Time stamp	Avg agreements per bin	Event Num	Features		
0:56-1:06	6	1	Very purposeful strikes	Perf action	Confidence
1:25-1:31	6	2	Swinging boxes, rhythm slow, looks in control	Control	
2:46-3:06	9	3	Air/box drumming in flow, begins to be varied	Complexity	
3:06-3:31	4.8	4	Air/box drumming in flow	Flow	
3:31-3:50	5.75	5	Brings in highest box, rhythm gets more complex	Addition	Rhythm
4:21-4:31	6.5	6	High partials start, rhythmic flow	New sound	Flow
4:51-5:01	14	7	Drumming on body/boxes in flow	Flow	Rhythm
5:06-5:11	6	8	Drumming on body/boxes in flow	Flow	Rhythm
5:16-5:26	6.5	9	One-handed drumming, in flow	Flow	Way of playing

Error events

Time stamp	Avg agreements per bin	Event Num	Features		
1:59-2:03	0.75	1	Rhythm seems inconsistent	Rhythm	Inconsistency
3:32	2	2	One box accidentally muted	Technical	

Laptop 1

Enjoyment events

Time stamp	Avg agreements per bin	Event Num	Features		
2:16-2:21	7	1	after some stopping, rhythm emerges	New	
4:06-4:11	8	2	There is an established rhythm	Established	

Error events

Time stamp	Avg agreements per bin	Event Num	Features		
0:08	2	1	Seems like false start	Intention	
0:49-0:52	0.75	2	Sounds don't seem intentional	Intention	
1:09-1:13	1	3	Silence, like something's not working	Intention	
1:26-1:29	1	4	Stopping and starting	Stop/start	Intention
1:34-1:37	0.75	5	Stopping and starting	Stop/start	Intention
1:39-1:41	1	6	Sounds come in then stop	Stop/start	Intention
2:01-2:04	0.75	7	Sounds come in then stop	Stop/start	Intention
4:00-4:02	1.33333333	8	Repeated sounds, not sure if intentional	Intention	

Audience 2

MOAI (Small)

Enjoyment events

Time stamp	Avg agreements per bin	Event Num	Features		
1:11-1:16	7	1	Repeating of rhythmic motif	Repetition	

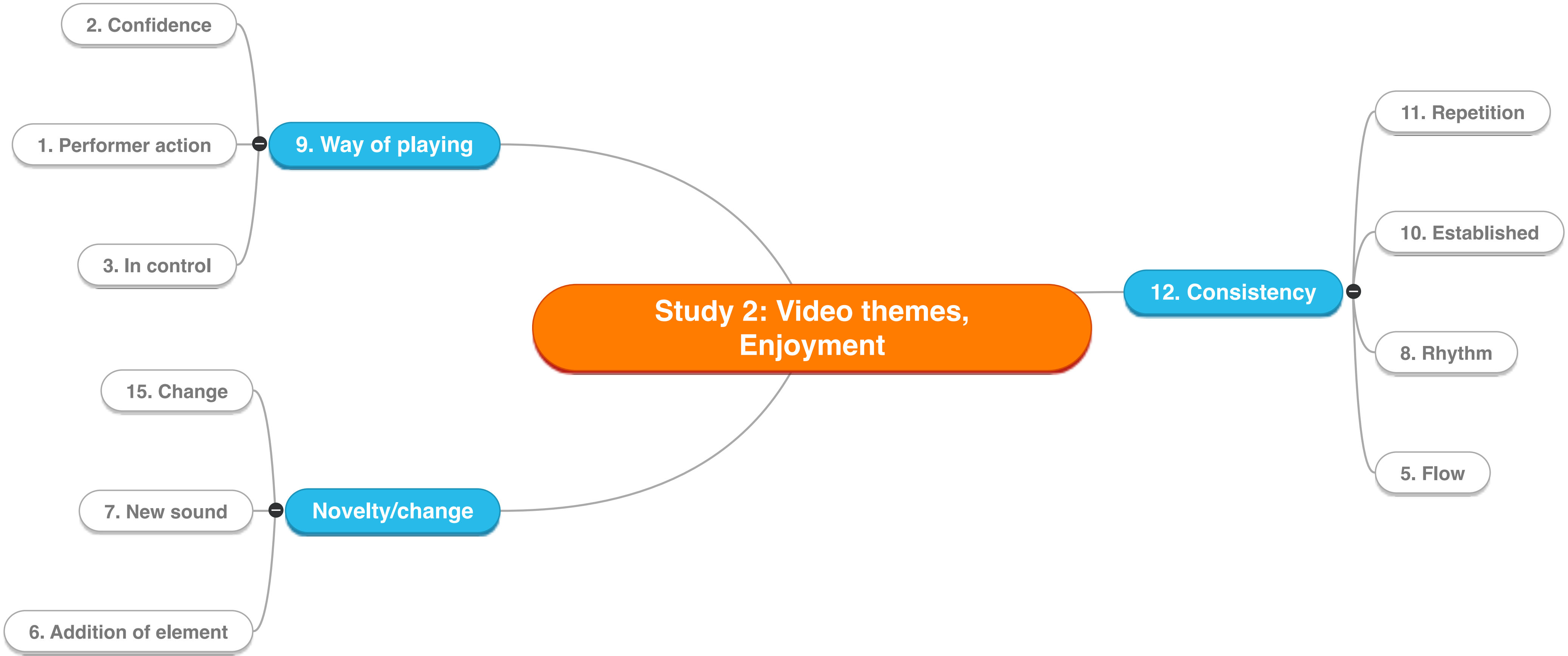
1:31-1:36	7	2	Repeating of rhythmic motif	Repetition		
4:01-4:06	6	3	Rhythm is consistent	Rhythm	consistency	
4:11-4:16	11	4	Rhythm flow, introduced a little more complexity	Rhythm	flow	new element
4:21-4:26	8	5	Rhythm flow	Rhythm	flow	
4:46-4:56	6	6	Rhythm flow	Rhythm	flow	
5:16-5:21	6	7	Box and body drumming, flow	Rhythm	flow	
6:01-6:06	8	8	Repeating of rhythmic motif	Repetition		
6:11-6:16	8	9	Repeating of rhythmic motif	Repetition		

Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
2:39-2:42	1	1	Intention	Intention		
3:53-3:54	1.5	2	?			
4:07-4:08	1	3	?			
5:02-5:05	1	4	?			

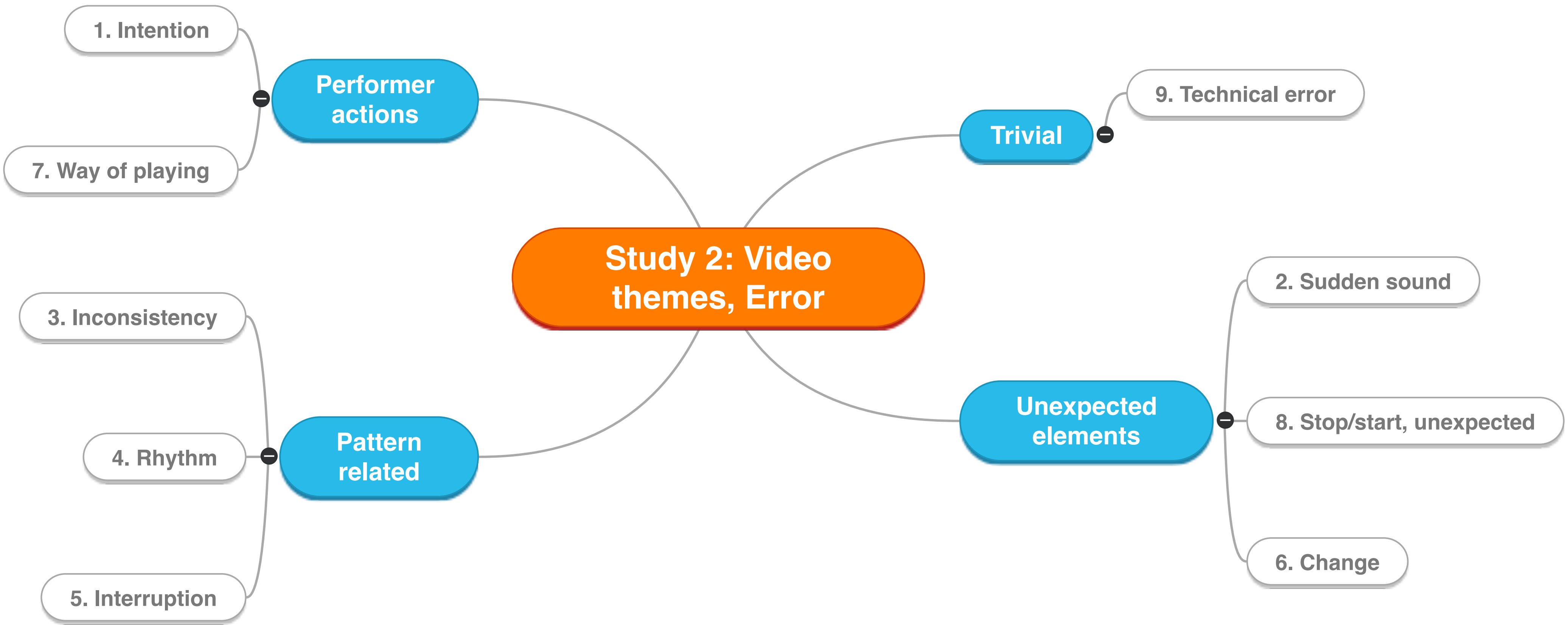
Laptop 2

Enjoyment events						
Time stamp	Avg agreements per bin	Event Num	Features			
1:30-1:35	6	1	Background rhythm	Rhythm		
3:20-3:25	6	2	Consistent rhythm with sound	Consistency		
3:45-3:50	6	3	Rhythmic flow	Flow		
4:30-4:45	12.33333333	4	Change to faster rhythm	Change		

Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:39	2	1	Stop, not sure if intentional	Rhythm interrupt	Intention	
1:12	2	2	?			
1:14-1:17	0.75	3	Stop, not sure if intentional	Rhythm interrupt	Intention	
2:46	2	4	Low sound, not sure if intentional	Sudden sound	Intention	
3:51	2	5	Rhythmic variation inconsistent	Rhythm	inconsistency	
4:20-4:23	0.75	6	Inconsistency in rhythm	Rhythm	inconsistency	
4:40-4:43	1	7	Drop in volume	Change		
4:44-4:47	0.75	8	She seems to be searching	Way of playing		
4:51-4:54	0.75	9	Rhythm inconsistent	Rhythm	inconsistency	
5:00-5:03	0.75	10	Rhythm inconsistent	Rhythm	inconsistency	
5:09	2	11	Sudden stop	Stop		
5:16-5:18	1	12	Short silence	Stop		



Study 2: Video themes, Enjoyment I disastrid I 2017-12-06 19:10:36 +0000


Study 2: Video themes, Error
Performer actions
1. Intention
7. Way of playing
Trivial
9. Technical error
Pattern related
3. Inconsistency
4. Rhythm
5. Interruption
Unexpected elements
2. Sudden sound
8. Stop/start, unexpected
6. Change

Appendix M

Study 3: Audience questionnaire

PERFORMANCE A: BEX

How much did you enjoy the performance? Please circle one.

1	2	3	4	5
Not at all				*Very much*

What did you like about the performance?

____

____

What did you dislike about the performance?

____

____

PERFORMANCE B: ROSIE

How much did you enjoy the performance? Please circle one.

1	2	3	4	5
Not at all				*Very much*

What did you like about the performance?

____

____

What did you dislike about the performance?

____

____

PERFORMANCE C: HENRIK

How much did you enjoy the performance? Please circle one.

1	2	3	4	5
Not at all				*Very much*

What did you like about the performance?

____

____

What did you dislike about the performance?

____

____

PERFORMANCE D: CALIE

How much did you enjoy the performance? Please circle one.

1	2	3	4	5
Not at all				*Very much*

What did you like about the performance?

____

What did you dislike about the performance?

____

PERFORMANCE E: LAUREL

How much did you enjoy the performance? Please circle one.

1	2	3	4	5
Not at all				*Very much*

What did you like about the performance?

____

What did you dislike about the performance?

____

PERFORMANCE F: ZANDS

How much did you enjoy the performance? Please circle one.

1	2	3	4	5
Not at all				*Very much*

What did you like about the performance?

____

What did you dislike about the performance?

____

END OF CONCERT SURVEY

1. What is your age? Please circle one:

<18   18-24   25-34   35-44   45-54   55-64   65+

2. How do you describe your gender? Please circle one:

Female   Male   Other   Prefer not to say

3. Are you involved in music in any of the following capacities (tick all that apply, and indicate your number of years of experience):

- Musician, ______ years
- Composer, _____ years
- Teacher, _____ years
- Student, _____ years
- Performer, ______ years
- Researcher, ______ years
- Hobbyist, _____ years
- Instrument designer, _____ years

4. Which musical genres do you listen to most frequently? Name up to 3:

____

5. Please rank the performances in order of preference, by placing numbers from 1 (your **favourite**) to 6 (your **least favourite**) beside each of the performances:

______ Performance A: Bex
______ Performance B: Rosie
______ Performance C: Henrik
______ Performance D: Calie
______ Performance E: Laurel
______ Performance F: Zands

6. Thinking about the performance you liked the most (rated 1), why was it your favourite?

____

____

7. Thinking about the performance you liked the least (rated 6), why was it your least favourite?

____

____

8. Did you notice errors during any of the performances? Please describe:

____

____

Finally ....

How would you describe your experience using the mobile phone interface?

____

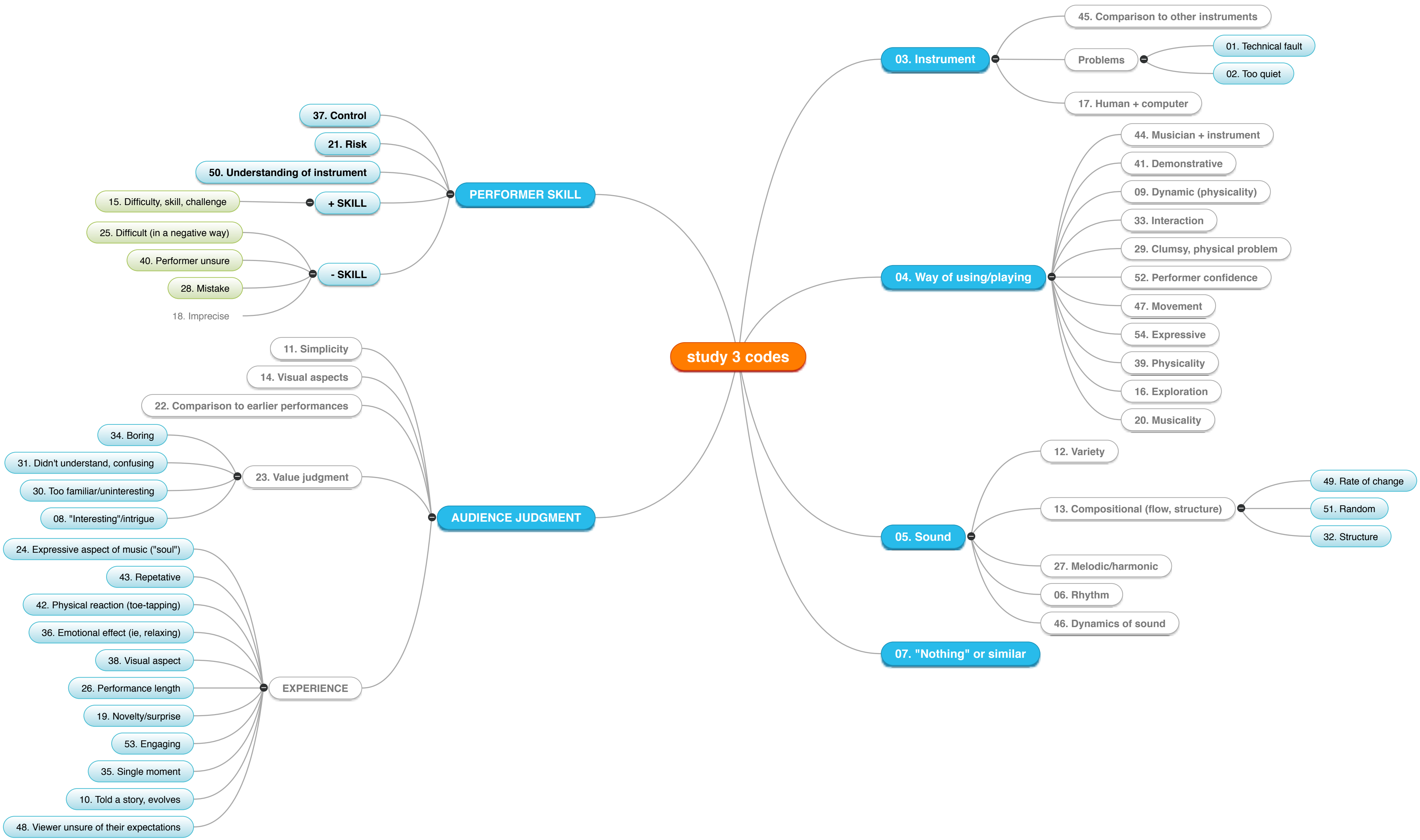
____

____

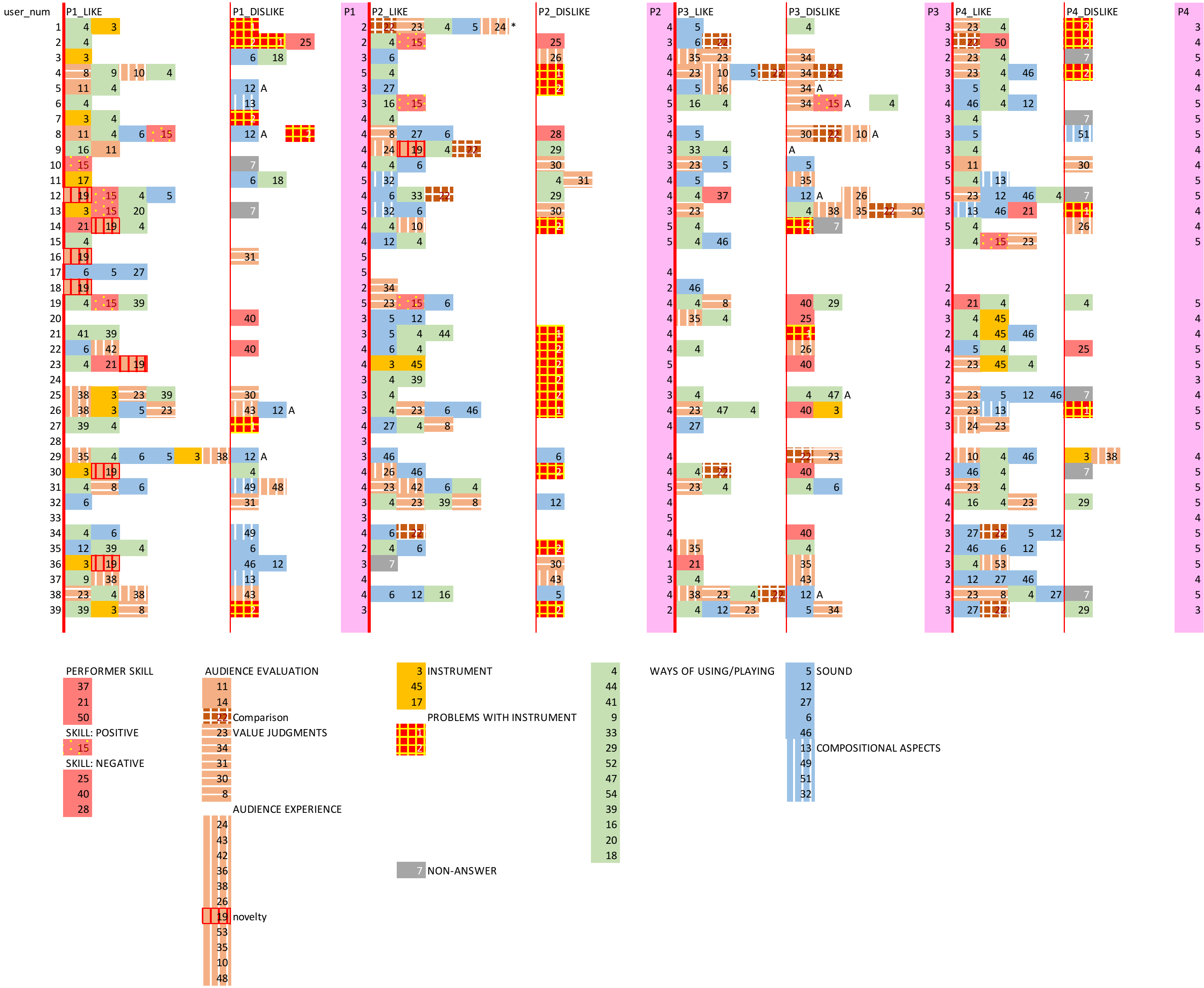
Thank you so much for your participation!

Appendix N

Study 3: Qualitative analysis



study 3 codes I disastrid I 2017-11-07 11:56:04 +0000



user_num
P1_LIKE
P1_DISLIKE
P1
P2_LIKE
P2_DISLIKE
P2
P3_LIKE
P3_DISLIKE
P3
P4_LIKE
P4_DISLIKE
P4
PERFORMER SKILL
SKILL: POSITIVE
SKILL: NEGATIVE
AUDIENCE EVALUATION
Comparison
VALUE JUDGMENTS
AUDIENCE EXPERIENCE
novelty
INSTRUMENT
PROBLEMS WITH INSTRUMENT
NON-ANSWER
WAYS OF USING/PLAYING
SOUND
COMPOSITIONAL ASPECTS

user_num	P5_LIKE	P5_DISLIKE	P5	P6_LIKE	P6_DILIKE	P6	P1	P2	P3	P4	P5	P6	confiden	way of	music/s	experien	skill	creativit	instrume	potentia	Expressi	Boring/e	Novelty	Visual	Risk
1	23 36	4	4	4 53		4	6	1							*1*	6			1	1		6			
2	19 22	25	4	53 4		4	6					1	6	6	6	*1*	1	1				1			
3	23 4 6	28	4	4 23	4 18	3			6	1					*1* 6										
4	8 5	2	4	4 54 15		5	1		6					6	*1*						1				
5	4		4	4 44		5	6					1		1 6				6					1		
6	4 19	6 18	2	3 4	54	4		1			6			1 6								6			
7	4		4	3 23		5		6				1	6		*1*									1	
8	22 46	4	4	23 4 15		5			6			1		1	*1* 6										
9	4 16		4	13		4					1							1		1					
10	21 13	7	5	4 11	7	5	6					1												1	
11	23 35	35	5	4 39	31 3	5																			
12	22 23 6	7	5	23 4	7	5		6				1			*1*										
13	15 4	18	5	55 4 44	5	4			6		1		6		*1*										1
14	52 4	7	5	6 35	7	5				6	1					*1*	6								
15	23 4 5 22		5	4	5 22	4					1	6			*1* 6								1		
16																									
17																									
18																									
19	23 35		4	4 8 39		5				6		1			*1*	6								1	
20	22 5 4			4 38 35		5			6			1		1 6											6
21	6 4		5	15 38 4		5			6		1				*1*			6							
22	5		3	39 4 49		4			1			6			*1*			6					1		
23	23 4	5		35	18	4			6	1			1 6											1 6	
24			2	4		4		1			6			1	6										
25	5 52	31	4	23 4	7	5			6			1		1										1 6	
26	5	1	3	4 42	7	5			6			1				*1*	6								
27	4	1	4			4																			
28											1			1											
29	23	18	3	3 10 38 4 35	18	4			6			1		1 6											
30	5 4	4	5	35 23 10	30 35	5			6		1			6	*1*										
31	4 23	25	4	4 38	31 5	5				1	6			1	*1* 6										
32	44 4 23	7	5	5 4	31 3	4	6				1			1 6			1 6								
33			5			4		1	6							*1* 6					1 6				
34	5 19		4	4	3	4			6	1				6			1								
35	54 21	45 5	5	6 12	18	4			6		1			1 6											
36	19 23		5	19 4		4		6		1				1 6											1 6
37		27 A	2	4 39		4	1		6													1 6			
38	23 5	12 A	3	23 4 15	7	4	6			1					*1*								1		
39	27 22	4 35	3		3	2				1		6		1 6											
Times rated highest:							6	3	15	2	3	3													
Times rated lowest:							2	4	1	7	9	11													

PERFORMER SKILL
37
21
50

SKILL: POSITIVE
15

SKILL: NEGATIVE
25
40
28

AUDIENCE EVALUATION
11
14
22 Comparison
23 VALUE JUDGMENTS
34
31
30
8

AUDIENCE EXPERIENCE
24
43
42
36
38
26
19 novelty
53
35
10
48

3 INSTRUMENT
45
17

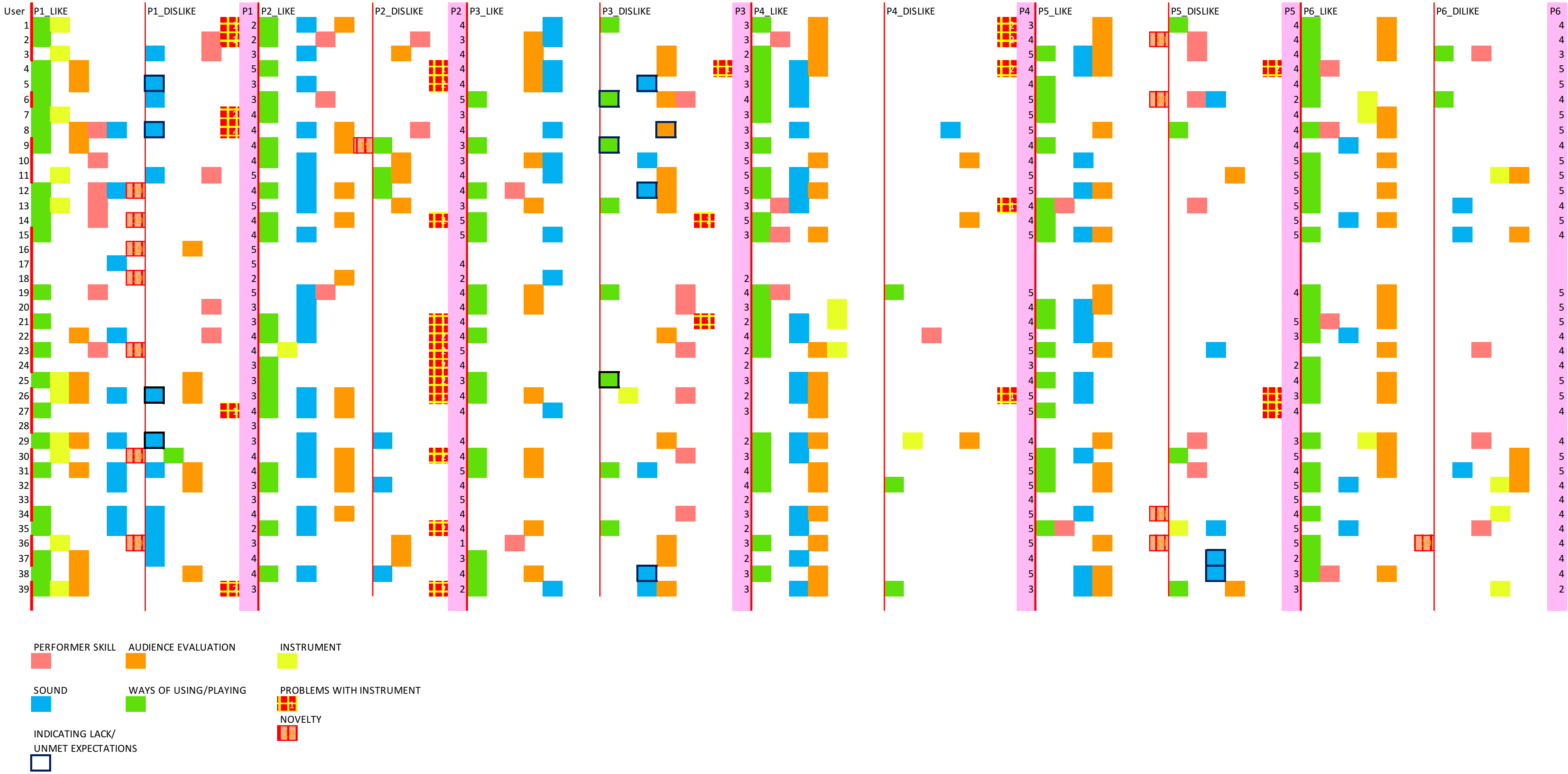
PROBLEMS WITH INSTRUMENT

7 NON-ANSWER

4 WAYS OF USING/PLAYIN
44
41
9
33
29
52
47
54
39
16
20
18

5 SOUND
12
27
6
46

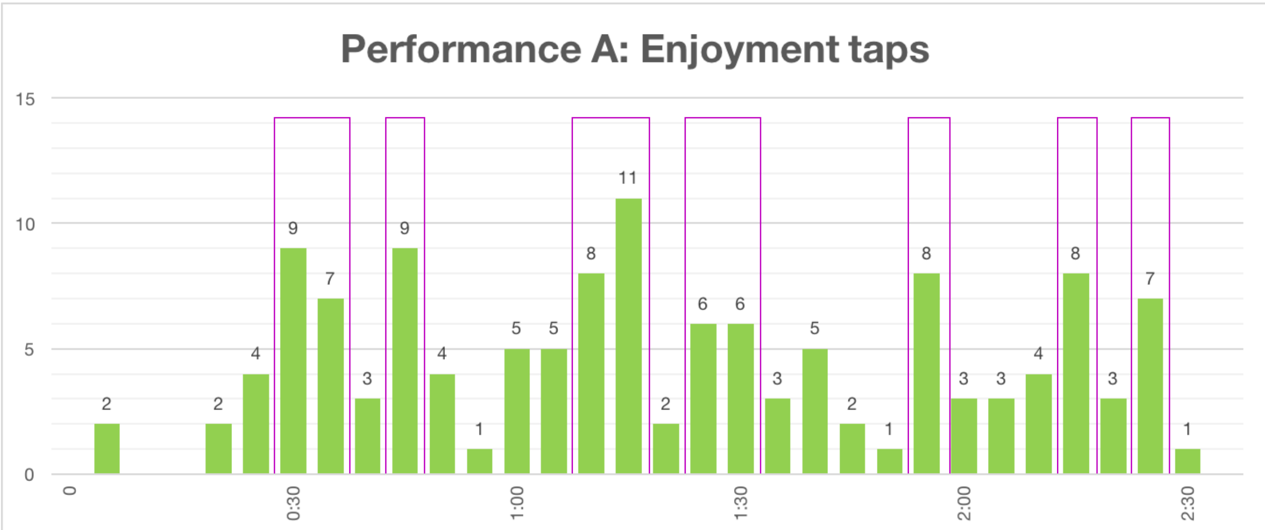
13 COMPOSITIONAL ASPECTS
49
51
32


User
P1_LIKE
P1_DISLIKE
P1
P2_LIKE
P2_DISLIKE
P2
P3_LIKE
P3_DISLIKE
P3
P4_LIKE
P4_DISLIKE
P4
P5_LIKE
P5_DISLIKE
P5
P6_LIKE
P6_DILIKE
P6
PERFORMER SKILL
AUDIENCE EVALUATION
INSTRUMENT
SOUND
WAYS OF USING/PLAYING
PROBLEMS WITH INSTRUMENT
NOVELTY
INDICATING LACK/
UNMET EXPECTATIONS

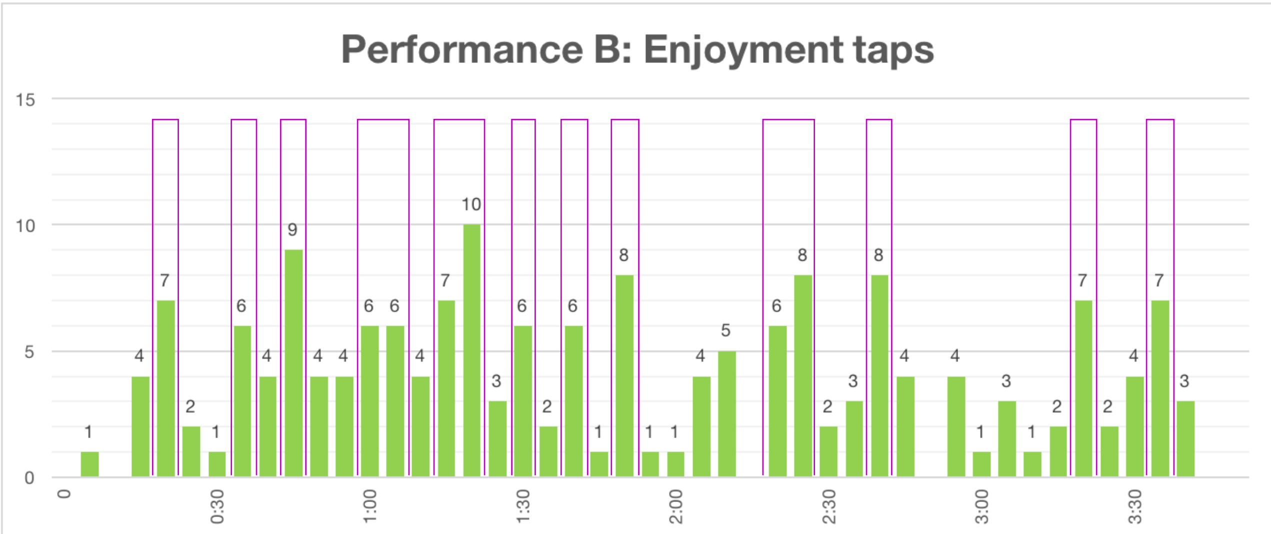
Appendix O

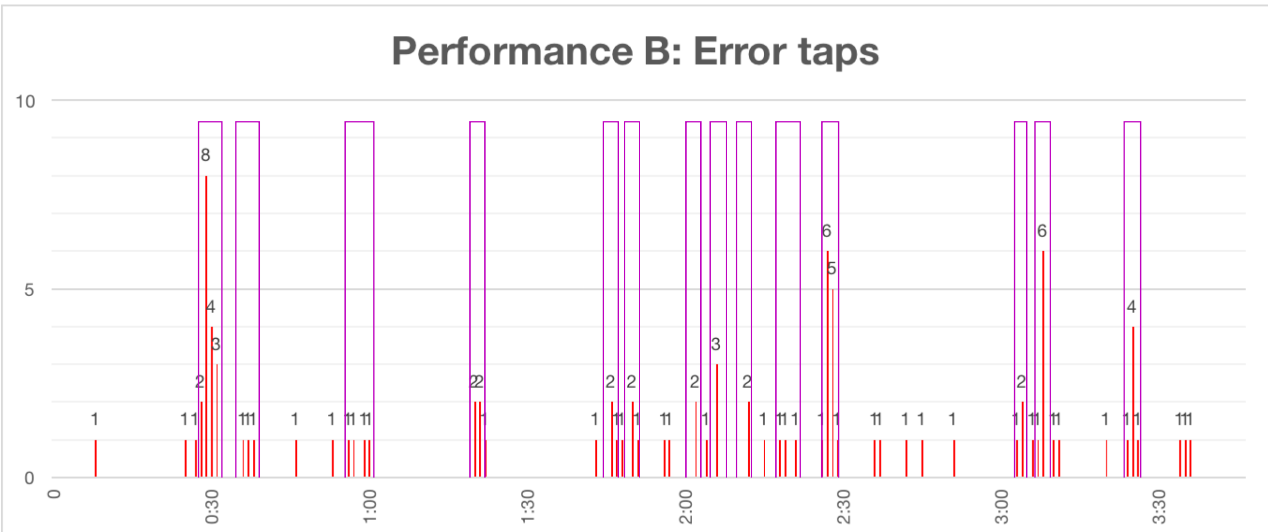
Study 3: Real-time analysis

Performance A
Real-time data visualisation

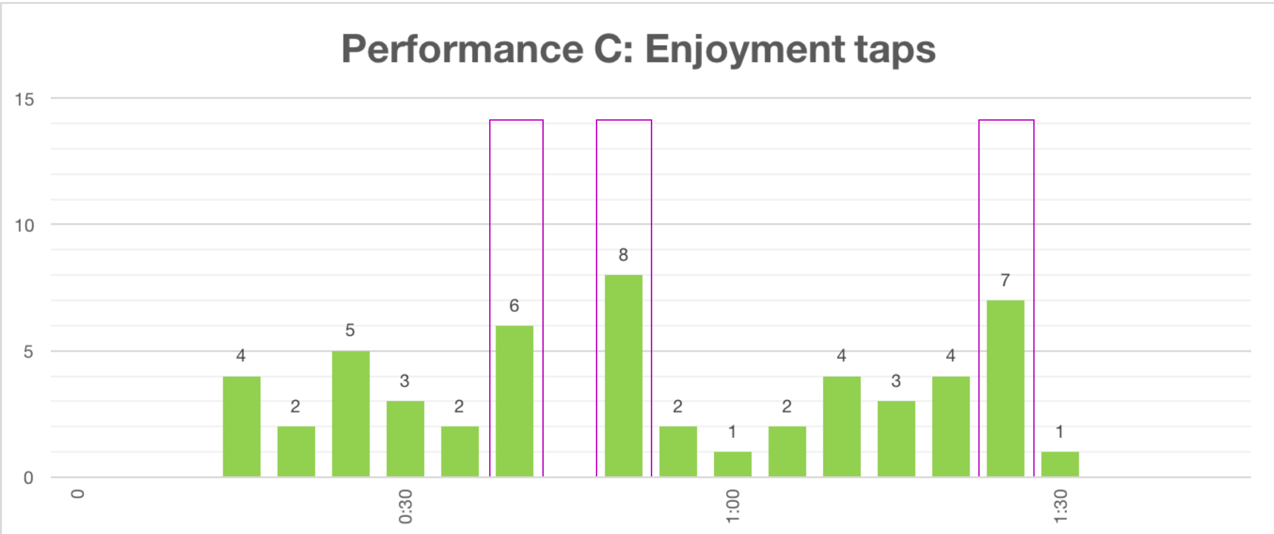


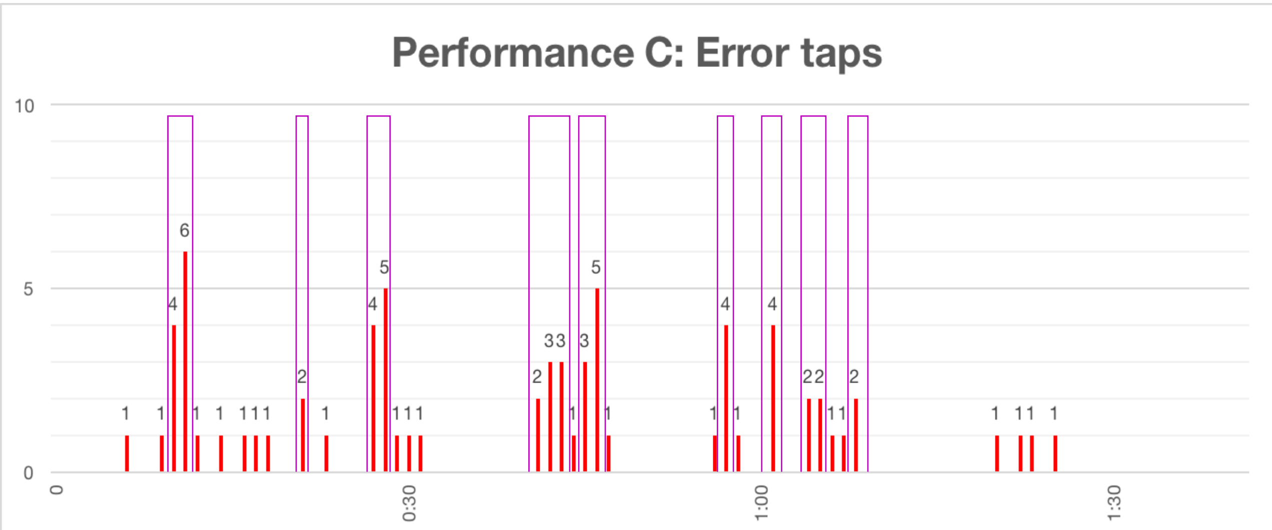
Performance B
Real-time data visualisation



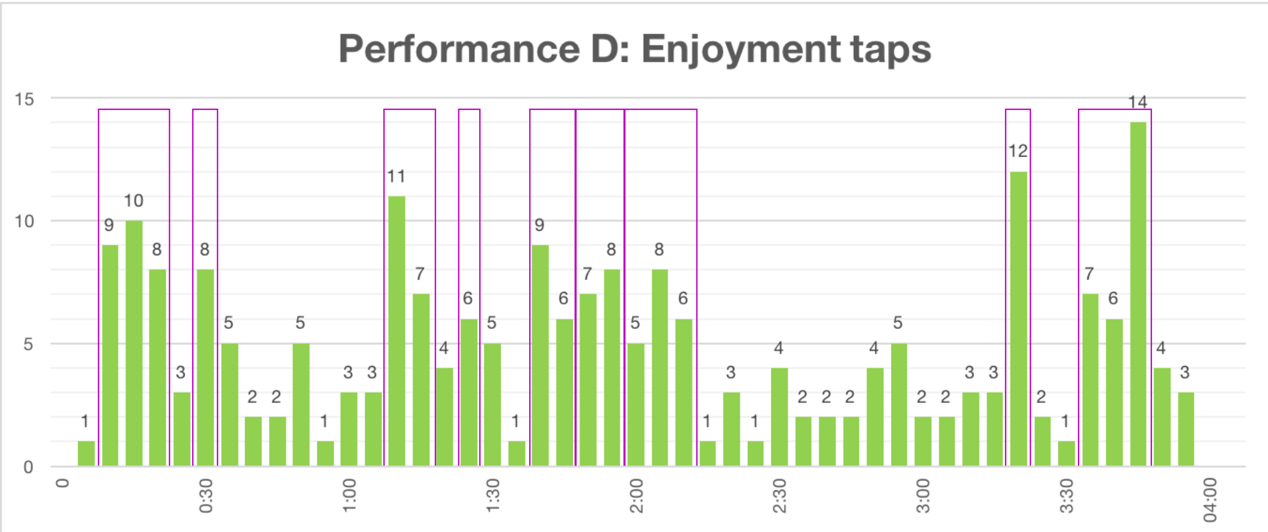


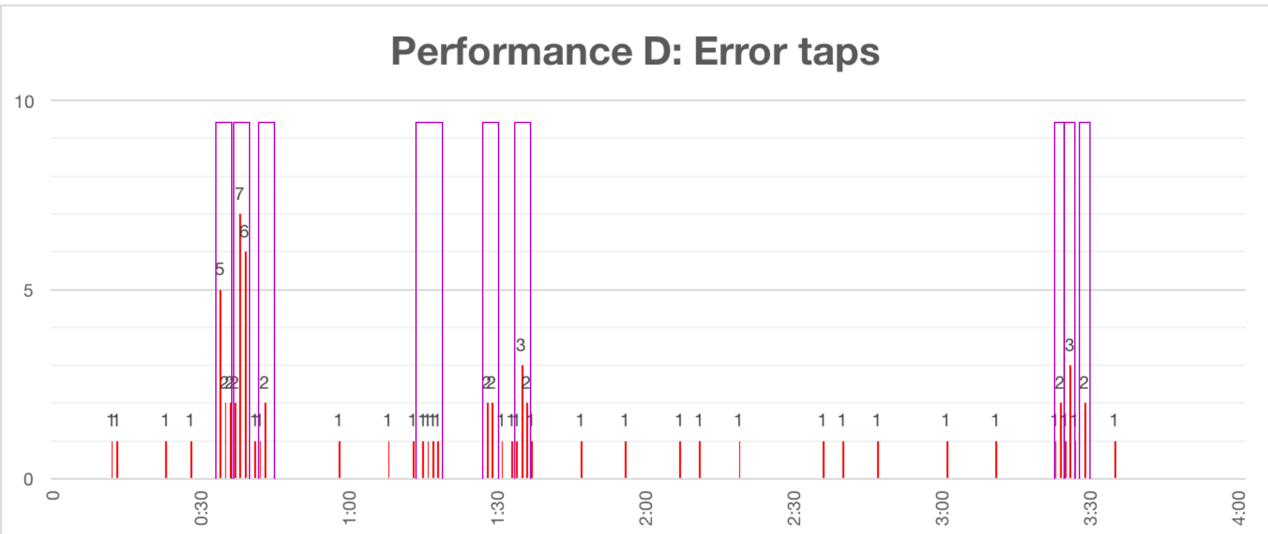
Performance C
Real-time data visualisation



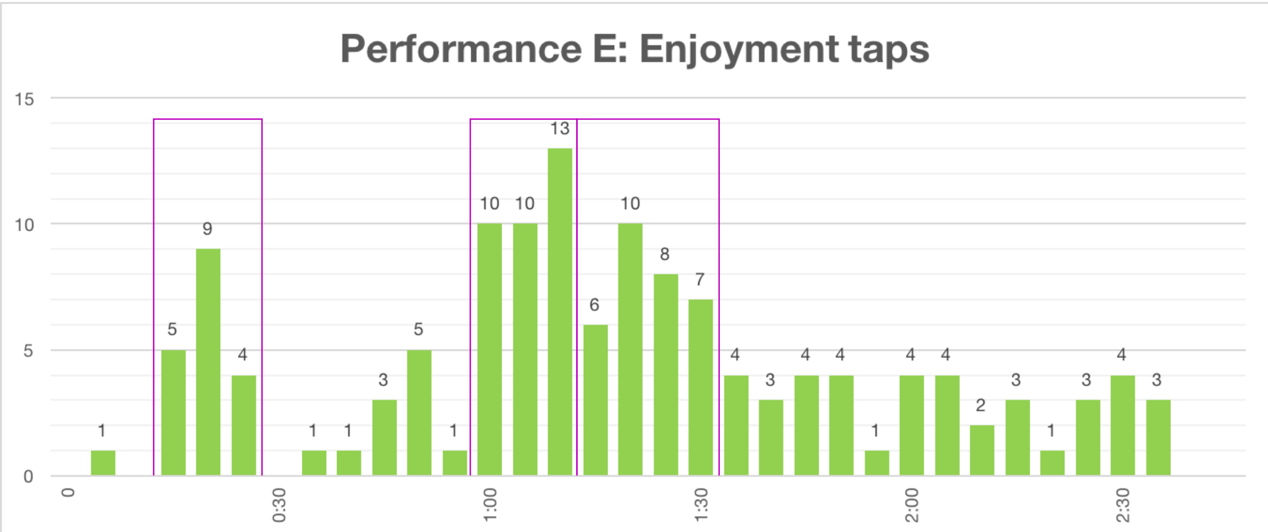


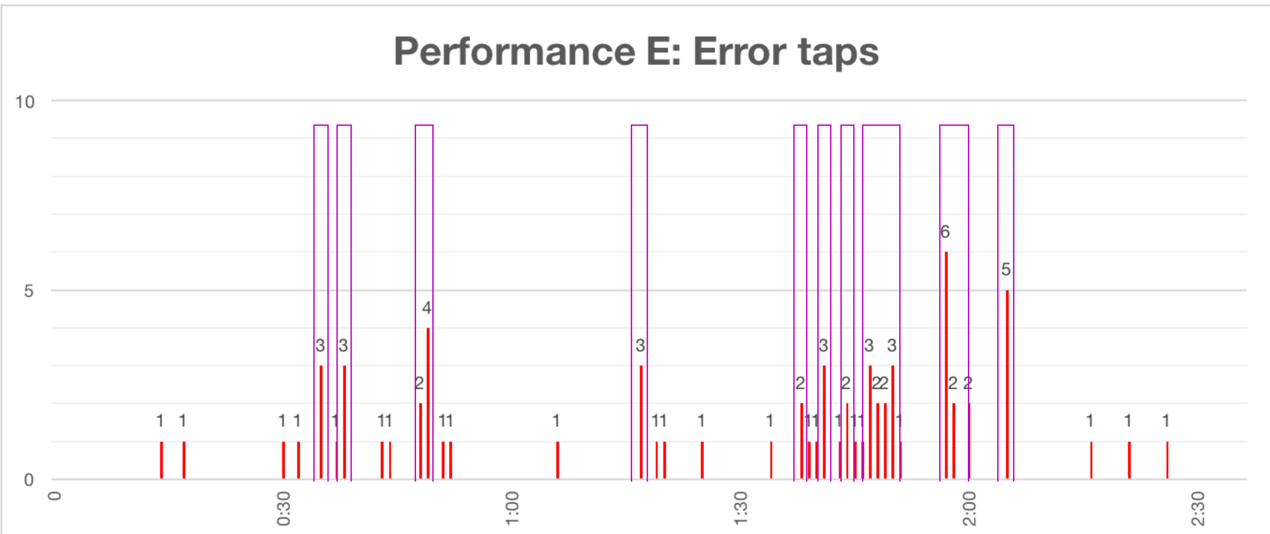
Performance D
Real-time data visualisation



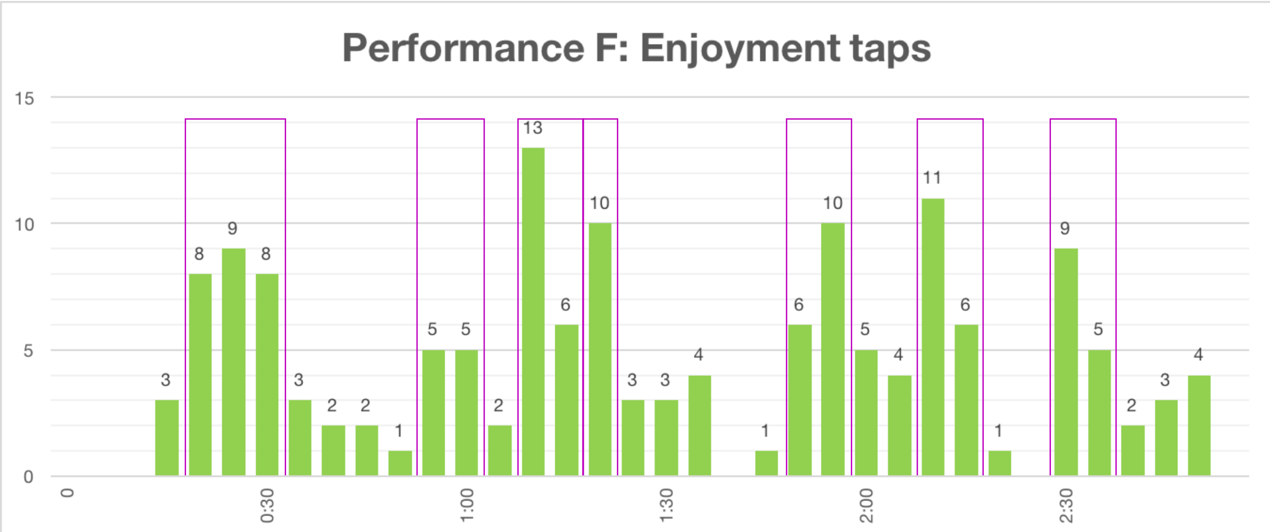


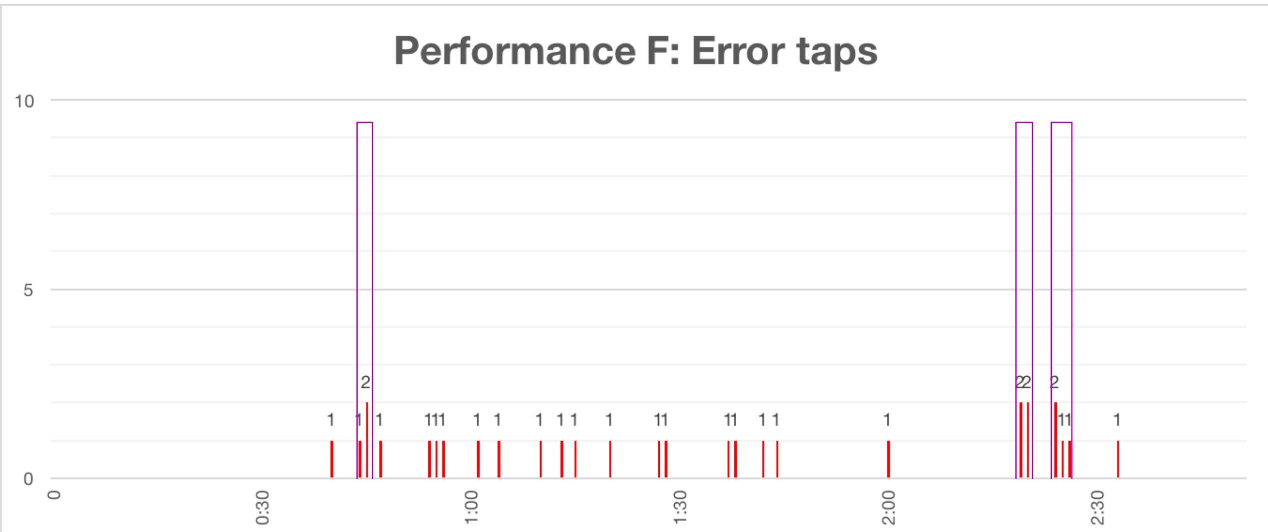
Performance E
Real-time data visualisation





Performance F
Real-time data visualisation





Appendix P

Study 3: Video documentation

Performance A

https://youtu.be/pdNrgpr7F2I

Performance B

https://youtu.be/iUV4GsXGmwk

Performance C

https://youtu.be/7WlvodWJXzM

Performance D

https://youtu.be/ve99LlGvBHA

Performance E

https://youtu.be/pylOOnsIHrY

Performance F

https://youtu.be/yZ1ukQLbW-w

Performance A

Enjoyment events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:30-0:40	8	1	Movement and on/off relationship are more obvious	Intention	Novelty	
0:45-0:50	9	2	Movement and on/off relationship are more obvious	Intention	Novelty	
1:10-1:20	9.5	3	Standing, establishes rhythm	Change	Establish	Physical
1:25-1:35	6	4	Change in handling, begins to turn instrument	Change		
1:55-2:00	8	5	Rhythm with feet is established	Rhythm	Establish	Technique
2:15-2:20	8	6	End	(End)		
2:25-2:30	7	7	After end of performance	(End)		

Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:07-0:09	3.666666667	1	Tapping, no sound (not moving it yet)	Intention		
0:12	4	2	Tapping, no sound (not moving it yet)	Intention		
0:45	2	3	?			
1:32	2	4	Didn't make sound on hit	Technical		
1:42	2	5	?			
1:44	4	6	Rhythm inconsistent	Inconsistent		
1:46	2	7	Rhythm inconsistent	Inconsistent		
1:58-2:00	5	8	Rhythm inconsistent	Inconsistent		
2:11	2	9	?			
2:18	5	10	She stops playing	Stop	Unexpected	

Performance B

Enjoyment events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:20-0:25	7	1	purposeful slap	Intention		
0:35-0:40	6	2	playing in lap, starts rhythm	New		
0:45-0:50	9	3	rhythm established	Establish		
1:00-1:10	6	4	Rhythm returns faster	Repetition	Technique	
1:15-1:25	13.5	5	Change handling to back and forth	Change		
1:30-1:35	6	6	Change to floor	Change		
1:40-1:45	6	7	Change in dynamics, quiet	Change		
1:50-1:55	8	8	Playing with feet	Novelty		
2:20-2:30	7	9	Lifts instrument, changes interaction	Change		
2:40-2:45	8	10	Rhythmic flow	Flow		
3:20-3:25	7	11	?			
3:35-3:40	7	12	End of performance	(End)		

Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:28-0:31	4.25	1	Adjusts feet and position	Adjust		
0:36-0:38	1	2	Unsure of tap	Unsure		
0:56-1:00	0.75	3	Break in rhythm	Break	Rhythm	

1:20-1:21	2	4	Instrument didn't repsond to hit	Technical		
1:46	2	5	Hesitates with foot, doesn't look purposeful	Hesitates		
1:50	2	6	Has to balance instrument	Struggle		
2:02	2	7	?			
2:06	3	8	Changes rhythm, doesn't seem intentional	Change	Intention	Rhythm
2:12	2	9	?			
2:18-2:21	0.75	10	Sloppy rhythm	Inconsistent	Rhythm	
2:27-2:28	5.5	11	Big gesture, no sound	Technical	Unmet expectation	
3:04	2	12	Break in rhythm flow	Break	Rhythm	
3:08	6	13	Gestures, no sound	Technical	Unmet expectation	
3:25	4	14	Gesture, no sound	Technical	Unmet expectation	

Performance C

Enjoyment events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:40-0:45	6	1	Gets into a rhythmical groove	Rhythm	Flow	
0:50-0:55	8	2	Groove	Rhythm	Flow	
1:25-1:30	7	3	End of performance	(End)		

Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:10-0:11	5.5	1	Rhythm inconsistent	Inconsistent	Rhythm	
0:21	2	2	Rhythm inconsistent	Inconsistent	Rhythm	
0:27-0:28	4.5	3	Tries to play with fingers, doesn't make sound	Technical	Unmet expectation	
0:41-0:43	2.666666667	4	Tempo inconsistent	Inconsistent	Rhythm	
0:45-0:46	4	5	Rhythm inconsistent	Inconsistent	Rhythm	
0:57	4	6	Rhythm inconsistent	Inconsistent	Rhythm	
1:01	4	7	Tries to play quietly, doesn't hit hard enough to sound	Technical		
1:04-1:06	2	8	Tries to play quietly, doesn't hit hard enough to sound	Technical		
1:08	2	9	Tries to play quietly, doesn't hit hard enough to sound	Technical		

Performance D

Enjoyment events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:10-0:25	9	1	First development of rhythm complexity	New		
0:30-0:35	8	2				
1:10-1:20	9	3	Consistent rhythm	Flow		
1:25-1:35	6	4	Starts varying the rhythm	Change		
1:40-1:50	7.5	5	Quick fingering, picks it up	Change	Technique	
1:50-2:00	7.5	6	Quick fingering, picks it up, tilts it	Change	Technique	
2:05-2:15	6.333333333	7	Plays by slapping	Change		
3:20-3:25	12	8	Quick tempos, lays it down	Change	Technique	
3:35-3:50	9	9	Quick consistent fingering	Technique		

Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:34-0:36	3	1	Rhythm inconsistent	Inconsistent		
0:37-0:39	5	2	Seems unsure	Unsure		
0:43	2	3	Seems unsure	Unsure		
1:15-1:17	1	4	Rhythm slows down, doesn't seem intentional	Intention		
1:28-1:29	2	5	Rhtyhm changes, doesn't seem intentional	Intention		
1:35-1:36	2.5	6	Rhythm slows down, doesn't seem intentional	Intention		
3:24	2	7	?			
3:26	3	8	Trying to put instrument on chair	Struggle		
3:29	2	9	?			

Performance E

Enjoyment events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:20-0:30	6	1	Slaps with shake, sits down, rhythm consistent	Change	Flow	
1:00-1:15	11	2	Starts knee rhythm, build up	Build up	Technique	
1:15-1:35	7.75	3	Fast knee-hand alternating, build up	Technique	Build up	

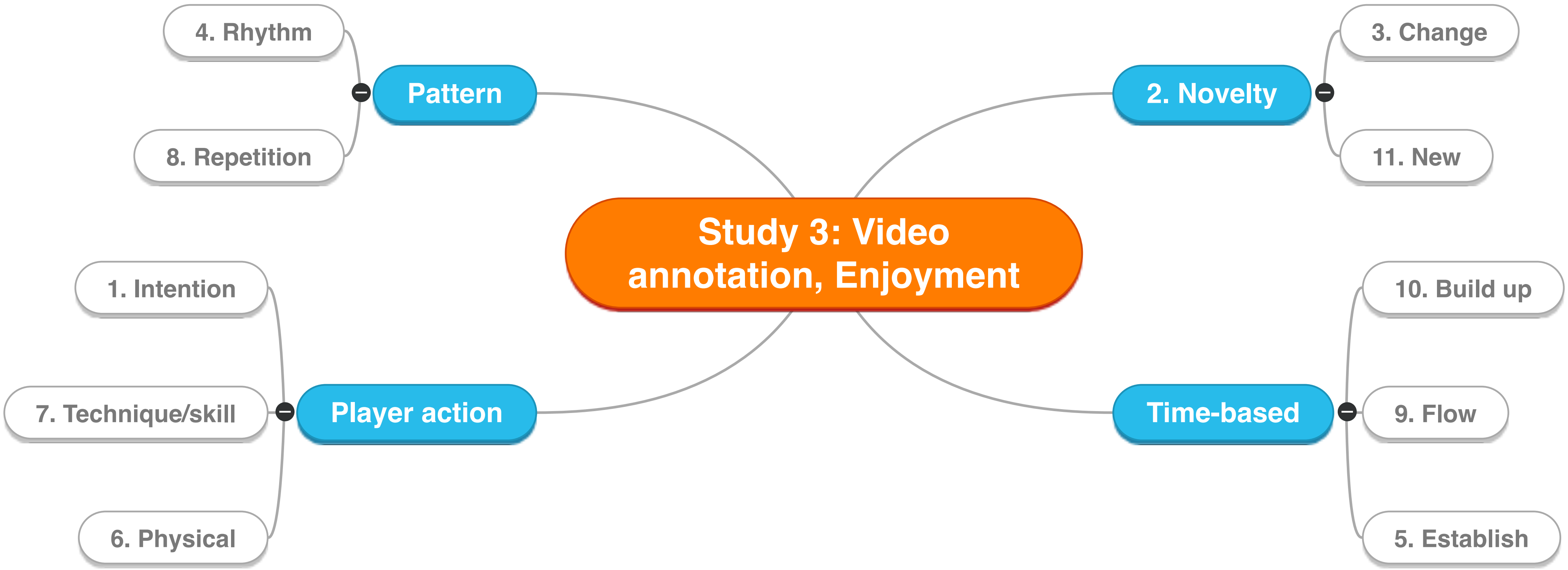
Error events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:35	3	1	Settling into chair	Adjustment		
0:37-0:38	2	2	Settling into chair	Adjustment		
0:48-0:49	3	3	Getting instrument on knee	Struggle		
1:17	3	4	Distortion in sound	Distortion		
1:38	2	5	Lots of distortion	Distortion		
1:41	3	6	Lots of distortion	Distortion		
1;44	2	7	Lots of distortion	Distortion		
1:47-1:50	2.5	8	Stops rhythm, instrument rolls around knee	Struggle	Break	
1:57-1:58	4	9	Lots of distortion	Distortion		
2:05	5	10	Inconsistent rhythm	Inconsistency		

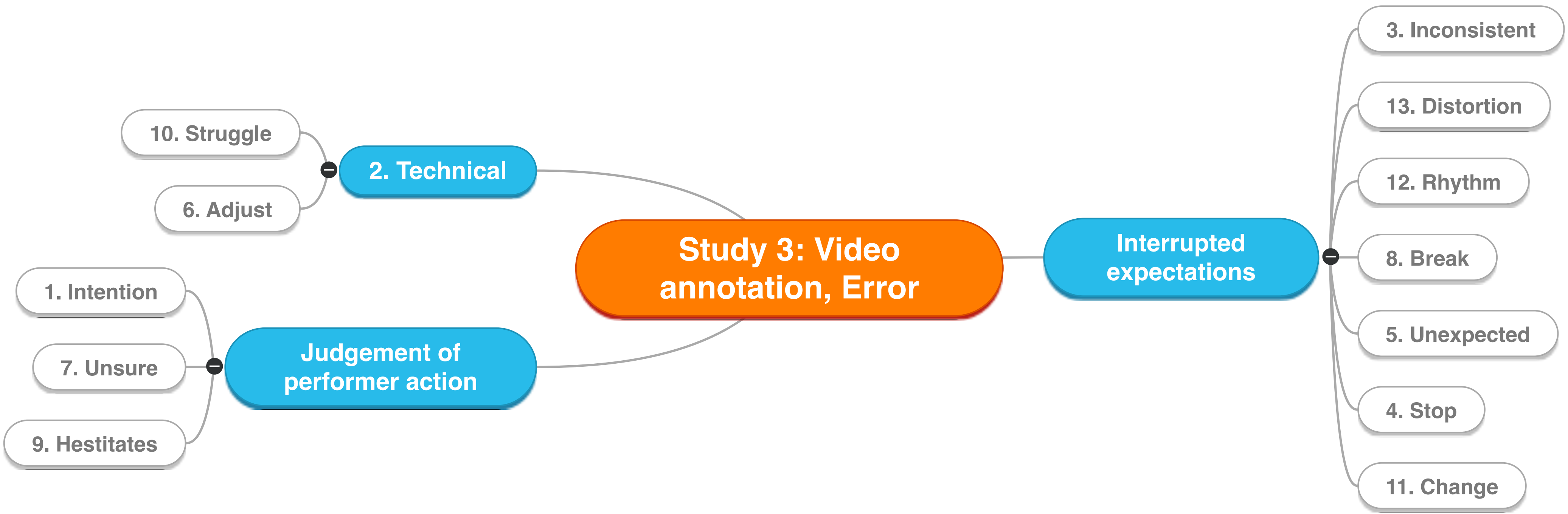
Performance F

Enjoyment events						
Time stamp	Avg agreements per bin	Event Num	Features			
0:20-0:35	8.333333333	1	Established rhythm	Establish		
0:55-1:05	5	2	Started tossing/playing, swinging	Change	Physical	Technique
1:10-1:20	9.5	3	Swinging in full effect	Change	Physical	Technique
1:20-1:25	10	4	Swinging in full effect	Change	Physical	Technique
1:50-2:00	8	5	Moved to toss-slapping, rhythm faster	Change	Physical	Technique
2:10-2:20	8.5	6	Very fast rhythm	Technique		
2:30-2:40	7	7	Toss and catch	Technique	Physical	End

Error events						

Time stamp	Avg agreements per bin	Event Num	Features			
0:44-0:45	2	1	Rhythm hiccup	Inconsistent		
2:19-2:20	2	2	?			
2:24-2:26	2	3	Catching on borders between electrodes, doesn't make s	Technical		


Study 3: Video annotation, Enjoyment
Pattern
4. Rhythm
8. Repetition
2. Novelty
3. Change
11. New
Player action
1. Intention
7. Technique/skill
6. Physical
Time-based
10. Build up
9. Flow
5. Establish



s03_videoThemes_err I disastrid I 2017-12-07

Bibliography

[1] Shore face detection engine. https://www.iis.fraunhofer.de/en/ff/sse/ils/tech/shore-facedetection.html. Accessed on 7 Jul 2017.

[2] Oxford english dictionary online, 2018. "risk", n. http://www.oed.com.ezproxy.library.qmul.ac.uk/view/Entry/166306?rskey=Fxys9G&result=1 (accessed Nov 11, 2017).

[3] Anshu Agarwal and Andrew Meyer. Beyond usability: evaluating emotional response as an integral part of the user experience. In *CHI Extended Abstracts*, pages 2919–2930. ACM, 2009.

[4] Phillip Alder. With a good fit, happily go higher. *Topeka Capital Journal*, 2012.

[5] Adam L Alter, Daniel M Oppenheimer, Nicholas Epley, and Rebecca N Eyre. Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136(4):569, 2007.

[6] Jack Armitage, Fabio Morreale, Andrew McPherson, et al. "the finer the musician, the smaller the details": Nimecraft under the microscope. In *Proceedings of the International Conference on New interfaces for Musical Expression*. ACM, 2017.

[7] Arubina. Moai set in the hillside at rano raraku, 2004. Via Wikimedia Commons: https://en.wikipedia.org/wiki/Moai#/media/File:Moai_Rano_raraku.jpg, accessed Aug 8 2017.

[8] Onyx Ashanti. This is beatjazz. Ted Foundation, 2011. Retrieved from: https://www.ted.com/talks/onyx_ashanti_this_is_beatjazz.

[9] P Auslander. *Liveness: Performance in a mediatized culture*. Routledge, 2008.

[10] Chris Baldick. indeterminacy.

[11] Jeronimo Barbosa, Joseph Malloch, Marcelo M Wanderley, and Stéphane Huot. What does "Evaluation" mean for the NIME community? In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 156–161. ACM, 2015.

[12] Louise Barkhuus and Tobias Jørgensen. *Engaging the crowd: studies of audience-performer interaction*. ACM, New York, USA, April 2008.

[13] Daniel Barolsky. Embracing imperfection in Benno Moiseiwitsch's prelude to Chopin. *Music Performance Research*, 2:48–60, 2008.

[14] Gregory Bateson. A theory of play and fantasy. In E Zimmerman and K Saien, editors, *The game design reader: A rules of play anthology*, pages 314–328. MIT Press, 2006.

[15] Jean Baudrillard. *Simulacra and simulation*. University of Michigan press, 1994.

[16] Jean Baudrillard. Art ... contemporary of itself. In Sylvère Lotringer, editor, *The Conspiracy of Art*, pages 89–97. Semiotext(e), 2005.

[17] M A Baytas, T Göksun, and O Özcan. The Perception of Live-sequenced Electronic Music via Hearing and Sight. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2016.

[18] Yardley Beers. *Introduction to the Theory of Error*. Addison-Wesley Publishing Company, Reading, USA, 2nd edition, 1962.

[19] Debra A Bekerian and John M Bowers. Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1):139, 1983.

[20] Victoria Bellotti, Maribeth Back, W Keith Edwards, Rebecca E Grinter, Austin Henderson, and Cristina Lopes. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, 2002. ACM.

[21] Steve Benford. Performing musical interaction: Lessons from the study of extended theatrical performances. *Computer Music Journal*, 34(4):49–61, 2010.

[22] Steve Benford, Holger Schnädelbach, Boriana Koleva, Rob Anastasi, Chris Greenhalgh, Tom Rodden, Jonathan Green, Ahmed Ghali, Tony Pridmore, Bill Gaver, Andy Boucher, Brendan Walker, Sarah Pennington, Albrecht Schmidt, Hans Gellersen, and Anthony Steed. Expected, sensed, and desired: A framework for designing sensing-based interaction. *Transactions on Computer-Human Interaction (TOCHI)*, 12(1):3–30, 2005.

[23] Florent Berthaut, David Coyle, James W Moore, and Hannah Limerick. Liveness Through the Lens of Agency and Causality. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2015.

[24] Florent Berthaut, Mark Marshall, Sriram Subramanian, and Martin Hachet. Rouages: Revealing the Mechanisms of Digital Musical Instruments to the Audience. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2013.

[25] Z Bilda. *Evaluating audience experience*. Creativity and cognition studios, January 2006.

[26] M Billinghurst and B Buxton. *Gesture based interaction*. Haptic input, 2011.

[27] David Birnbaum, Rebecca Fiebrink, Joseph Malloch, and Marcelo M Wanderley. Towards a dimension space for musical devices. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 192–195. ACM, 2005.

[28] Elizabeth Ligon Bjork and Robert A Bjork. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2:59–68, 2011.

[29] Jennifer Blackwell. Music Education, Recording Technology, And The Illusion Of Perfection. *GEMS (Gender, Education, Music, and Society), the on-line journal of GRIME (Gender Research in Music Education)*, 2013.

[30] Susanne Bødker. *When second wave HCI meets third wave challenges.* ACM, New York, New York, USA, October 2006.

[31] Till Bovermann, Alberto de Campo, Hauke Egermann, Sarah-Indriyati Hardjowirogo, and Stefan Weinzierl. *Musical Instruments in the 21st Century: Identities, Configurations, Practices.* Springer, 2016.

[32] O Bown, R Bell, and A Parkinson. Examining the perception of liveness and activity in laptop music: Listeners' inference about what the performer is doing from the audio alone. In *Proceedings of the Conference on New Interfaces for Musical Expression.* ACM, 2014.

[33] O Bown, Lian Loke, Sam Ferguson, and Dagmar Reinhardt. Distributed Interactive Audio Devices: Creative strategies and audience responses to novel musical interaction scenarios. In *Proceedings of the Conference on New Interfaces for Musical Expression*, 2014.

[34] Brandtzaeg, P.; Følstad, A; Heim, J. Lessons from enjoyment. In P C Wright, K Overbeeke, A F Monk, and M A Blythe, editors, *Funology: from usability to enjoyment*, chapter 5. Kluwer Academic Publishers, 2003.

[35] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[36] W Brooker and D Jermyn. *The audience studies reader.* Psychology Press, 2003.

[37] P Bull and M Noordhuizen. The mistiming of applause in political speeches. *Journal of Language and social psychology*, 19(3):275–294, 2000.

[38] David Byrne. *How Music Works*. McSweeney's, San Francisco, 2012.

[39] Claude Cadoz. Supra-instrumental interactions and gestures. *Journal of New Music Research*, 38(3):215–230, 2009.

[40] Claude Cadoz and Marcelo M Wanderley. Gesture-music, 2000.

[41] John Cage. *Silence*. Lectures and Writings. Marion Boyars, London, 21 edition, August 2009.

[42] A Camurri, S Hashimoto, M Ricchetti, and A Ricci. Eyesweb: Toward gesture and affect recognition in interactive dance and music systems. *Computer Music Journal*, 24(1):57–69, 2000.

[43] Antonio Camurri, Giovanni De Poli, Marc Leman, and Gualtiero Volpe. A multi-layered conceptual framework for expressive gesture applications. In *Proceedings of the International MOSART Conference*, 2001.

[44] Ian Carr. *Keith Jarrett: The man and his music*. Da Capo Press, 1992.

[45] Kim Cascone. The aesthetics of failure:"post-digital" tendencies in contemporary computer music. *Computer Music Journal*, 24(4):12–18, 2000.

[46] Anton P. Chekhov. *The three sisters*. Dover Publications, 1993.

[47] Perry Cook. Principles for designing computer music controllers. In *Proceedings of the International Conference on New interfaces for Musical Expression*, pages 1–4. ACM, 2001.

[48] Mihaly Csikszentmihalyi and Isabella Selega Csikszentmihalyi. *Optimal experience: Psychological studies of flow in consciousness*. Cambridge university press, 1992.

[49] Mihaly Csikszentmihalyi and Rick Emery Robinson. *The art of seeing: An interpretation of the aesthetic encounter*. Getty Publications, 1990.

[50] Meagan E Curtis and Jamshed J Bharucha. Memory and musical expectation for tones in cultural context. *Music Perception: An Interdisciplinary Journal*, 26(4):365–375, 2009.

[51] Luke Dahl and Ge Wang. Sound bounce: Physical metaphors in designing mobile music performance. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 178–181. ACM, 2010.

[52] F. Davis. *In the Moment: Jazz in the 1980s*. Oxford University Press, 1986.

[53] Tom Davis. Towards a Relational Understanding of the Performance Ecosystem. *Organised Sound*, 16(2):120–124, 2011.

[54] Brenda Dervin. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2):36–46, December 1998.

[55] Brenda Dervin. Chaos, order and sense-making: A proposed theory for information design. *Information design*, pages 35–57, 1999.

[56] J D'Escrivan. To sing the body electric: Instruments and effort in the performance of electronic music. *Contemporary Music Review*, pages 1–10, 2006.

[57] P Desmet. Measuring emotion: Development and application of an instrument to measure emotional responses to products. In M A Blythe, A F Monk, K Overbeeke, and P C Wright, editors, *Funology: From Usability to Enjoyment*, pages 111–123. Funology, 2005.

[58] Diana Deutsch. *Psychology of Music*. Academic Press, Amsterdam, third edition edition, August 2013.

[59] J Diaz. The Fate of Auto-Tune. *Music and Technology*, 2009.

[60] Connor Diemand-Yauman, Daniel M Oppenheimer, and Erikka B Vaughan. Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, 118(1):111–115, 2011.

[61] Alan Dix. Designing for Appropriation. In *Proceedings of the British HCI Group Annual Conference on People and Computers*, pages 27–30. British Computing Society, 2007.

[62] Christopher Dobrian and Daniel Koppelman. The 'E' in NIME: Musical expression with new computer interfaces. In *Proceedings of the*

Conference on New Interfaces for Musical Expression, pages 277–282. ACM, 2006.

[63] Benjamin Duvall. Ex-easter island head: Six sticks. `https://youtu.be/sjKt2kQ_TQw?t=3m7s`, 2016. [YouTube; accessed 9-Sept-2016].

[64] Peter Elsdon. Style and the Improvised in Keith Jarrett's Solo Concerts. *Jazz Perspectives*, 2(1):51–67, 2008.

[65] G Emerson and H Egermann. Mapping, Causality and the Perception of Instrumentality: Theoretical and Empirical Approaches to the Audience's Experience of Digital Musical Instruments. *Musical Instruments in the 21st Century*, 2017.

[66] Gyorgy Fazekas, Mathieu Barthet, and Mark B Sandler. Mood conductor: emotion-driven interactive music performance. In *Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 726–726. IEEE, 2013.

[67] S Fels. Designing for intimacy: Creating new interfaces for musical expression. In *Proceedings of the IEEE*, pages 672–685. IEEE, 2004.

[68] Sidney Fels, Ashley Gadd, and Axel Mulder. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound*, 7(2):109–126, 2002.

[69] R Fiebrink, D Trueman, N C Britt, and M Nagai. Toward Understanding Human-Computer Interaction In Composing The Instrument. In *Proceedings of the International Computer Music Conference*, 2010.

[70] Jason Freeman and Mark Godfrey. Technology, real-time notation, and audience participation in flock. In *ICMC*, 2008.

[71] A C Fyans, Michael Gurevich, and Paul Stapleton. Examining the spectator experience. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 451–454. ACM, 2010.

[72] A Cavan Fyans and Michael Gurevich. Perceptions of skill in performances with acoustic and electronic instruments. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 495–498. ACM, 2011.

[73] A. Cavan Fyans, Michael Gurevich, and Paul Stapleton. Spectator understanding of error in performance. In *CHI Extended Abstracts*, New York, USA, 2009. ACM.

[74] A Cavan Fyans, Michael Gurevich, and Paul Stapleton. Where did it all go wrong? a model of error from the spectator's perspective. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 171–172. ACM, 2009.

[75] A Cavan Fyans, Michael Gurevich, and Paul Stapleton. Examining the spectator experience. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 1–4. ACM, 2010.

[76] A Gabrielsson and P N Juslin. Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of music*, 1996.

[77] William W Gaver, Jacob Beaver, and Steve Benford. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 233–240. ACM, 2003.

[78] Lalya Gaye, Lars Erik Holmquist, Frauke Behrendt, and Atau Tanaka. Mobile music technology: Report on an emerging community. In *Proceedings of the 2006 Conference on New interfaces for musical expression*, pages 22–25. IRCAM—Centre Pompidou, 2006.

[79] Alfred Gell. The technology of enchantment and the enchantment of technology. *Anthropology, art and aesthetics*, pages 40–63, 1992.

[80] Gilbert G Germain. *Spirits in the Material World: The Challenge of Technology*. Rowman & Littlefield, 2009.

[81] James J Gibson. *The ecological approach to visual perception: Classic edition*. Psychology Press, 2014.

[82] Rolf Inge Godøy. Gestural affordances of musical sound. *Musical gestures: Sound, movement, and meaning*, pages 103–125, 2010.

[83] L Goehr. *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Clarendon Press, 1992.

[84] M. Gurevich, P. Stapleton, and A. Marquez-Borbon. Style and constraint in electronic musical instruments. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2010.

[85] Michael Gurevich. Interacting with Cage: Realising classic electronic works with contemporary technologies. *Organised Sound*, 20(03):290–299, 2015.

[86] Michael Gurevich and A Cavan Fyans. Digital musical interactions: Performer–system relationships and their perception by spectators. *Organised Sound*, 16(2):166–175, 2011.

[87] Michael Gurevich and A Cavan Fyans. Digital Musical Interactions: Performer/system relationships and their perception by spectators. *Organised Sound*, 16(2):166–175, 2011.

[88] Michael Gurevich, Paul Stapleton, and Peter D Bennett. Designing for style in new musical interactions. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 213–217. ACM, 2009.

[89] Michael Gurevich and Jeffrey Treviño. Expression and its discontents: toward an ecology of musical creation. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 106–111. ACM, 2007.

[90] A Hamilton. The art of improvisation and the aesthetics of imperfection. *British Journal of Aesthetics*, 2000.

[91] Tim Harford. How messy problems can inspire creativity. Ted Foundation, 2015. Retrieved from: `https://www.ted.com/talks/tim_harford_how_messy_problems_can_inspire_creativity/`.

[92] Steve Harrison, Deborah Tatar, and Phoebe Sengers. The three paradigms of HCI. *Alt-CHI Session at the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–18, 2007.

[93] Ian Hattwick and Marcelo M Wanderley. A dimension space for evaluating collaborative musical performance systems. In *Proceedings of the Conference on New Interfaces for Musical Expression*, volume 12, pages 21–23. ACM, 2012.

[94] Kate Hayes, Mathieu Barthet, Yongmeng Wu, Leshao Zhang, and Nick Bryan-Kinns. A participatory live music performance with the open symphony system. In *CHI Extended Abstracts*, pages 313–316. ACM, 2016.

[95] Robert Henke. The Uncanny Valley. I spent five years developing and refining my real time laser drawing software for Lumière. Now the results are as complex as technology allows for. `http://www.facebook.com`, May 2017. [Facebook post; accessed 21-July-2017].

[96] Amelie Hinrichsen and Till Bovermann. Post-dmi musical instruments. In *Proceedings of the Audio Mostly Conference*, pages 124–131. ACM, 2016.

[97] Thom Holmes. *Electronic and Experimental Music*. Pioneers in Technology and Composition. Psychology Press, 2002.

[98] J Hook, G Schofield, R Taylor, and T Bartindale. Exploring HCI's relationship with liveness. *CHI Extended Abstracts*, 2012.

[99] William T Hsu and Marc H Sosnick. Evaluating interactive music systems: An hci approach. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 25–28. ACM, 2009.

[100] Jennifer Huang and Carol Lynne Krumhansl. What does seeing the performer add? it depends on musical style, amount of stage behavior, and audience expertise. *Musicae Scientiae*, page 1029864911414172, 2011.

[101] D B Huron. Sweet anticipation: Music and the psychology of expectation, 2006.

[102] Jonathan Impett. Interaction, simulation and invention: a model for interactive music. In *Proceedings of ALMMA 2001 Workshop on Artificial Models for Musical Applications*, pages 108–119. Cosenza, Italy, 2001.

[103] Robert H Jack, Tony Stockman, and Andrew McPherson. Rich gesture, reduced control: the influence of constrained mappings on performance technique. In *Proceedings of the International Conference on Movement in Computing*, 2017.

[104] Keith Jarrett. The köln concert, 1975. Decca Records.

[105] A R Jensenius. To gesture or not? An analysis of terminology in NIME proceedings 2001-2013. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2014.

[106] Carey Jewitt, Jeff Bezemer, and Kay O'Halloran. *Introducing multimodality*. Routledge, 2016.

[107] S. Jordà. Instruments and players: Some thoughts on digital lutherie. *J. New Music Research*, 33:321–341, 2004.

[108] S. Jorda and S. Mealla. A methodological framework for teaching, evaluating and informing NIME design with a focus on expressiveness and mapping. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2014.

[109] Sergi Jorda. Instruments and Players: Some Thoughts on Digital Lutherie. *Journal of New Music Research*, 2004.

[110] S Jordà. FMOL: Toward User-Friendly, Sophisticated New Musical Instruments. *Computer Music Journal*, 26(3):23–39, 2014.

[111] P N Juslin. Five facets of musical expression: A psychologist's perspective on music performance. *Psychology of music*, 31(3):273–302, 2003.

[112] B Kane. *Sound Unseen: Acousmatic Sound in Theory and Practice*. Oxford University Press, 2014.

[113] Stanley Kaplan and B John Garrick. On the quantitative definition of risk. *Risk analysis*, 1(1):11–27, 1981.

[114] K Katevas, PGT Healey, and M T Harris. Robot stand-up: engineering a comic performance. In *Proceedings of the Workshop on Humanoid Robots and Creativity at the IEEE-RAS International Conference on Humanoid Robots Humanoids*. IEEE, 2014.

[115] Miguel Pina e Cunha Ken N Kamoche and Joao Vieira da Cunha edt. Introduction and Overview. In Miguel Pina e Cunha, Joao Vieira Da Cunha, and Ken N Kamoche, editors, *Organisational Improvisation*, pages 1–12. Psychology Press, 2002.

[116] Sarah Kettley. More art than science? craft as creative methodology. In *Proceedings of the Art and Sound Symposium*, 2017.

[117] Hazrat Inayat Khan. *The mysticism of sound and music*. Shambhala Publications, 1996.

[118] G Klein, B M Moon, and R R Hoffman. Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems*, 2006.

[119] Michael Klingbeil. Spear homepage. `http://www.klingbeil.com/spear/`.

[120] Michael Klingbeil. Software for spectral analysis, editing, and synthesis. In *ICMC*, 2005.

[121] R Kostelanetz. Conversing with cage, 2003.

[122] Justin Kruger, Derrick Wirtz, Leaf Van Boven, and T William Altermatt. The effort heuristic. *Journal of Experimental Social Psychology*, 40(1):91–98, 2004.

[123] Gordon Kurtenbach and Eric A Hulteen. Gestures in human-computer communication. In Brenda Laurel and S Joy Mountford, editors, *The art of human-computer interface design*. Addison Wesley, 1990.

[124] C H Lai and T Bovermann. Audience Experience in Sound Performance. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2013.

[125] Celine Latulipe, Erin A Carroll, and Danielle Lottridge. *Love, hate, arousal and engagement: exploring audience responses to performing arts*. exploring audience responses to performing arts. ACM, New York, New York, USA, May 2011.

[126] Marco Lehmann and Reinhard Kopiez. The influence of on-stage behavior on the subjective evaluation of rock guitar performances. *Musicae Scientiae*, 17(4):472–494, 2013.

[127] Marc Leman. *Embodied music cognition and mediation technology*. Mit Press, 2008.

[128] Raymond Loewy. *Never Leave Well Enough Alone*. Simon & Schuster, 1951.

[129] E F Loftus and J C Palmer. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, 1974.

[130] Sylvère Lotringer and Paul Virilio. The accident of art. 2005.

[131] Thor Magnusson. Designing constraints: Composing and performing with digital musical systems. *Computer Music Journal*, 34(4):62–73, 2010.

[132] Thor Magnusson. Interfacing sound: visual representation of sound in musical software instruments. In *Musical Instruments in the 21st Century*, pages 153–166. Springer, 2017.

[133] Thor Magnusson. Musical organics: a heterarchical approach to digital organology. *Journal of New Music Research*, 46(3):286–303, 2017.

[134] Adnan Marquez-Borbon, Michael Gurevich, A C Fyans, and Paul Stapleton. Designing digital musical interactions in experimental contexts. *contexts*, (June):373–376, 2011.

[135] M T Marshall, P Bennett, M Fraser, and S Subramanian. Emotional response as a measure of liveness in new musical instrument performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012.

[136] Andrew McPherson. The magnetic resonator piano: Electronic augmentation of an acoustic grand piano. *Journal of New Music Research*, 39(3):189–202, 2010.

[137] Andrew McPherson. Bela: An embedded platform for low-latency feedback control of sound. *The Journal of the Acoustical Society of America*, 141(5):3618–3618, 2017.

[138] Carolina Brum Medeiros and Marcelo M Wanderley. A comprehensive review of sensors and instrumentation methods in devices for musical expression. *Sensors*, 14(8):13556–13591, 2014.

[139] Rosa Menkman. *The Glitch Moment(um)*. Institute of Network Cultures, Amsterdam, 2011.

[140] W J Millard. A history of handsets for direct measurement of audience response. *International Journal of Public Opinion Research*, 1992.

[141] Eduardo Reck Miranda and Marcelo M Wanderley. *New digital musical instruments: control and interaction beyond the keyboard*, volume 21. AR Editions, Inc., 2006.

[142] Andrew Monk, Marc Hassenzahl, Mark Blythe, and Darren Reed. *Funology: designing enjoyment*. designing enjoyment. ACM, New York, New York, USA, April 2002.

[143] F R Moore. The dysfunctions of MIDI. *Computer Music Journal*, 12(1), 1988.

[144] F Morreale and A P McPherson. Design for longevity: Ongoing use of instruments from nime 2010-14. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2017.

[145] T. Murray-Browne, D. Mainstone, N. Bryan-Kinns, and M. Plumbley. The medium is the message: Composing instruments and performing mappings. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2011.

[146] Tim Murray-Browne. *Interactive music: Balancing creative freedom with musical development*. PhD thesis, Queen Mary University of London, 2012.

[147] Eugene Narmour. The Top-down and Bottom-up Systems of Musical Implication: Building on Meyer's Theory of Emotional Syntax. *Music Perception: An Interdisciplinary Journal*, 9(1):1–26, 1991.

[148] Eugene Narmour. *The analysis and cognition of melodic complexity: The implication-realization model*. University of Chicago Press, 1992.

[149] NASA. One small step. `https://www.nasa.gov/62284main_onesmall2.wav`, 1969. [Accessed 30-Jun-2016].

[150] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 249–256. ACM, 1990.

[151] Donald A Norman. *Emotional design: Why we love (or hate) everyday things*. Basic Civitas Books, 2004.

[152] Donald A Norman. *The Design of Everyday Things*. MIT Press, Cambridge, MA, 2013.

[153] H L O'Brien and E G Toms. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the Association for Information Science and Technology*, 2008.

[154] S O'Modhrain. A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal*, 35(1):28–42, 2011.

[155] B Ostertag. *Why computer music sucks*. Resonance, 1998.

[156] Dan Overholt. The overtone violin. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 34–37. ACM, 2005.

[157] G Paine. Towards unified design guidelines for new interfaces for musical expression. *Organised Sound*, 14:142–155, 2009.

[158] G Paine. New musical instrument design considerations. *IEEE Multimedia*, 20(4):76–84, 2013.

[159] Marcus T Pearce and Geraint A Wiggins. Expectation in melody: The influence of context and learning. *Music Perception: An Interdisciplinary Journal*, 23(5):377–405, 2006.

[160] O Perrotin and C d'Alessandro. Visualizing Gestures in the Control of a Digital Musical Instrument. In *Proceedings of the Conference on New Interfaces for Musical Expression*. ACM, 2014.

[161] Richard A Peterson. In search of authenticity. *Journal of Management Studies*, 42(5):1083–1098, 2005.

[162] P Q Pfordresher, C Palmer, and M K Jungers. Speed, accuracy, and serial order in sequence production. *Cognitive science*, 2007.

[163] Peggy Phelan. *Unmarked.* The Politics of Performance. Routledge, July 1993.

[164] Cornelius Poepel. On interface expressivity: a player-based study. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 228–231. ACM, 2005.

[165] Ezra Pound. Make it new. In *Literary Essays of Ezra Pound.* New Directions Publishing, 1968.

[166] I. Poupyrev, M. J. Lyons, S. Fels, and T. Blaine Bean. *New interfaces for musical expression.* ACM, 2001.

[167] Jens Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3):257–266, 1983.

[168] James Reason. Human error: models and management. *BMJ: British Medical Journal*, 320(7237):768–770, 2000.

[169] S Reeves, S Benford, C O'Malley, and M Fraser. Designing the spectator experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2005.

[170] EAVI Goldsmiths. Best of NIME 2014. `https://www.youtube.com/watch?v=beLxqGKvI-M`, 2014. [YouTube; accessed 22-Nov-2017].

[171] Music Technology QCGU. NIME 2016 'Bio-Vortex: Exploring Wet Machines As Interfaces For New Musical Expression' Ben Freeth and John Bowers, 2016. Retrieved from: `https://www.youtube.com/watch?v=beLxqGKvI-M`.

[172] A Rosenblueth, N Wiener, and J Bigelow. Behavior, purpose and teleology. *Philosophy of science*, 10(1), 1943.

[173] Daniel M Russell, Peter Pirolli, George Furnas, Stuart K Card, and Mark Stefik. *Sensemaking workshop CHI 2009.* ACM, New York, New York, USA, April 2009.

[174] Luigi Russolo. The Art of Noises: Futurist Manifesto. In *Audio Culture: Readings in modern music*, pages 10–14. Continuum International, New York, 2010.

[175] W A Schloss. Using contemporary technology in live performance: The dilemma of the performer. *Journal of New Music Research*, 2003.

[176] Carl Emil Seashore. *Objective analysis of musical performance*, volume 4. The University Press, 1936.

[177] Phoebe Sengers and Bill Gaver. Staying open to interpretation: engaging multiple meanings in design and evaluation. In *the 6th ACM conference*, pages 99–108, New York, New York, USA, June 2006. ACM.

[178] C E Shannon. A mathematical theory of communication. *SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[179] Claude Elwood Shannon. Communication in the presence of noise. In *Proceedings of the IRE*, pages 10–21. IEEE, 1949.

[180] J G Sheridan and N Bryan-Kinns. Designing for performative tangible interaction. *International Journal of Arts and Technology*, 2008.

[181] Jennifer G Sheridan, Nick Bryan-Kinns, and Alice Bayliss. Encouraging witting participation and performance in digital live art. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it*, volume 1, pages 13–23. British Computer Society, 2007.

[182] Jennifer Gayle Sheridan. *Digital Live Art: mediating wittingness in playful arenas*. PhD thesis, University of Lancaster, 2006.

[183] C Stevens, R Glass, E Schubert, and J Chen. Methods for measuring audience reactions. In *Proceedings of the Inaugural International Conference on Music Communication Science*, 2007.

[184] John Storey. *Cultural theory and popular culture: A reader*. University of Georgia Press, 2006.

[185] Dan Stowell, Andrew Robertson, Nick Bryan-Kinns, and Mark D Plumbley. Evaluation of live human–computer music-making: Quantitative and qualitative approaches. *International Journal of Human-Computer Studies*, 67(11):960–975, 2009.

[186] Louis H Sullivan. The tall office building artistically considered. *Lippincotts Magazine*, March 1896. Retrieved from: `https://archive.org/details/tallofficebuildi00sull`.

[187] Jesse Sykes, Stephen O'Malley, Greg Anderson, Wata, Atsuo, and Takeshi. The sinking belle. In *Altar*. 2006. [YouTube; accessed 6-Jul-2016].

[188] A Tanaka. Musical performance practice on sensor-based instruments. *Trends in Gestural Control of Music*, 2000.

[189] Atau Tanaka. Interaction, Experience and the Future of Music. In *Consuming Music Together*, pages 267–288. Springer, Dordrecht, Berlin/Heidelberg, 2006.

[190] Atau Tanaka. Sensor-based musical instruments and interactive music. chapter 12, pages 233–257. 2009.

[191] Atau Tanaka. Biomuse to bondage: Corporeal interaction in performance and exhibition. In *Intimacy Across Visceral and Digital Performance*, pages 159–169. Springer, 2012.

[192] Debora V Thompson and Elise Chandon Ince. When disfluency signals competence: The effect of processing difficulty on perceptions of service agents. *Journal of Marketing Research*, 50(2):228–240, 2013.

[193] W F Thompson, P Graham, and F A Russo. Seeing music performance: Visual influences on perception and experience. *Semiotica*, 2005.

[194] Chia-Jung Tsay. Sight over sound in the judgment of music performance. In *Proceedings of the National Academy of Sciences*, volume 110, pages 14580–14585. National Academy Sciences, 2013.

[195] Sherry Turkle and Seymour Papert. Epistemological pluralism: Styles and voices within the computer culture. *Signs: Journal of women in culture and society*, 16(1):128–157, 1990.

[196] Kai Tuuri, Jaana Parviainen, and Antti Pirhonen. Who controls who? embodied control within human–technology choreographies. *Interacting with Computers*, pages 1–18, 2017.

[197] J Tyrangiel. Auto-tune: Why pop music sounds perfect. *Time Magazine*, 2009.

[198] Doug Van Nort and Nicolas Castagné. Mapping in digital musical instruments. In *Enaction and enactive interfaces: A handbook of terms*, pages 191–192. Enactive Systems Books, 2007.

[199] Dianne Verdonk. Visible excitation methods: energy and expressiveness in electronic music performance. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 42–43. ACM, 2015.

[200] Paul Virilio. Speed and information: Cyberspace alarm! *ctheory*, pages 8–27, 1995.

[201] Paul Virilio. *Politics of the Very Worst: An Interview by Philip Pettit.* Semiotext (e), 1999.

[202] F Visi, E Coorevits, R Schramm, and E R Miranda. Musical Instruments, Body Movement, Space, and Motion Data: Music as an Emergent Multimodal Choreography. *Human Technology*, 13(1):58–81, 2017.

[203] Kendall L Walton. How Marvelous! Toward a Theory of Aesthetic Value. *The Journal of Aesthetics and Art Criticism*, 51(3):499–510, 1993.

[204] 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.

[205] Chen Wang, Erik N Geelhoed, Phil P Stenton, and Pablo Cesar. Sensing a live audience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1909–1912. ACM, 2014.

[206] Simon Waters. Performance Ecosystems: Ecological approaches to musical interaction. *EMS: Electroacoustic Music Studies Network*, pages 1–20, 2007.

[207] C Wen-Chung. Open rather than bounded. *Perspectives of New Music*, 5(1), 1966.

[208] L Foreman Wernet and B Dervin. Everyday Encounters with Art: Comparing Expert and Novice Experiences. *Curator: The Museum Journal*, 2016.

[209] D Wessel and M Wright. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 2002.

[210] Virginia Woolf. *Mr. Bennet and Mrs. Brown.* Virginia Woolf, 2017.

[211] Peter Wright, Mark Blythe, and John McCarthy. User experience and the idea of design in hci. In *International Workshop on Design, Specification, and Verification of Interactive Systems*, pages 1–14. Springer, 2005.

[212] Gareth W Young, David Murphy, and Jeffrey Weeter. A functional analysis of haptic feedback in digital musical instrument interactions. In *Musical Haptics*, pages 95–122. Springer, 2018.

[213] Victor Zappi, Andrew Allen, and Sidney Fels. The Hyper Drumhead: Making Music with a Massive Real-time Physical Model. In *International Computer Music Conference*, 2017.

[214] Victor Zappi and Andrew McPherson. Dimensionality and Appropriation in Digital Musical Instrument Design. In *Proceedings of the Conference on New Interfaces for Musical Expression*, pages 455–460. ACM, 2014.