The Design of a Smartphone-based Digital Musical Instrument for Jamming

Benjamin Swift

October 25, 2012

A thesis submitted for the degree of
Doctor of Philosophy of
The Australian National University

Institutional Address

Research School of Computer Science
CS&IT Building 108
Australian National University
Canberra 0200
AUSTRALIA

Supervisory Panel

Assoc. Prof. Henry Gardner (Panel Chair)
Research School of Computer Science
Australian National University

Dr. Alistair Riddell
School of Art
Australian National University

Dr. David Bissell
Research School of the Social Sciences
Australian National University

Written in Emacs with the help of AUCTeX.
Set in Linux Libertine with the help of KOMA-Script and XeLaTeX.

For Joy

Except where otherwise indicated, this thesis is my own original work.

Ben Swift
October 25, 2012

Publications

During my PhD programme I have published five double blind peer-reviewed publications, two of which were single-author. I was also a musical contributor to the (peer reviewed) Winter 2011 Computer Music Journal DVD.

Long Paper	B. Swift (2012a). "Becoming-Sound: Affect and Assemblage in Improvisational Digital Music-Making". In: *CHI '12: Proceedings of the International Conference on Human Factors in Computing Systems.*
Book Chapter	B. Swift (2012b). "Chasing a Feeling: Experience in Computer Supported Jamming". In: *Music and Human-Computer Interaction.* Springer
Music Composition	B. Swift (2011). "Impish Grooves". In: *Computer Music Journal* 35.4, pp. 119–137
Workshop Paper	B. Swift, H. J. Gardner, and A. Riddell (2011). "A Chasing After the Wind: Experience in Computer-Supported Group Music-Making". In: *When Words Fail: What can music interaction tell us about HCI? Workshop at BCS HCI 2011*
Long Paper	B. Swift, H. J. Gardner, and A. Riddell (2010). "Engagement Networks in Social Music-making". In: *OZCHI '10: Proceedings of the Australasian Conference on Computer-Human Interaction*
Long Paper	B. Swift, H. J. Gardner, and A. Riddell (2009). "Distributed Performance in Live Coding". In: *ACMC '09: Proceedings of the Australasian Computer Music Conference*

Acknowledgements

Many people have helped me during this long journey—far too many to list. Still, there are a few folks who deserve special mention.

To Joy, who at the beginning was a friend and now is my fiancé, thankyou. Your support and love has been such a blessing, especially as my insecurity climaxed in these last few months. I love you. To Mum and Dad, too, I am so lucky to have parents as supportive as you guys, and I love you lots. To the rest of my family, thanks also for your love and support. I'm a lucky bloke to have you guys around.

To Torb, Jaebong, Henry, Alistair, Dave and Andrew, with whom I have shared the academic journey of the last few years, thanks for everything. I'm pretty sure I would have gone crazy if I had done this on my own, and you guys are a big part of the reason why that (mostly) didn't happen. Henry, you talked me off the ledge more than once, and your expertise and humility as a researcher are an inspiration to me. Alistair, you were always around to provide your expert perspective as an artist and a reminder to keep doing things that were creatively satisfying. And Dave, you generously answered my email out of the blue and gave me a crash course on geography, providing a valuable 'outsider' perspective on all this technology stuff. I am so grateful to you all for your help and advice.

To my housemate Dave, John-boy, and all my mates who saw the thesis writing process up close, thanks for your patience. I hope it was interesting to see how the sausage was made, and I trust you're all suitably inoculated against any desire to give it a go yourself. To *all* my friends, who all supported me in their own ways, thanks. I'm super grateful to have so many loved ones in my life.

There are several cafés in and around the ANU which deserve a mention as well. The staff at Biginelli Espresso, the Street Theatre, and Tilleys have seen countless hours of my head poking out over my laptop as I hacked away at this thing, particularly towards the end of the thesis writing process. Thanks for that.

Finally, in the spirit of J. S. Bach, *soli Deo gloria.* I can do everything through him who gives me strength.

Abstract

Open-ended human-computer interactions, such as those in interactive digital art and music, are an increasingly popular area of study in Human-Computer Interaction (HCI). They provide an opportunity to examine playfulness, creativity and expression and challenge conventional HCI notions of quality, evaluation and how to measure success.

Jamming—improvisational group music making—is often held up as an example of open-ended creativity. This thesis describes the development of *Viscotheque*, an iPhone-based Digital Musical Instrument (DMI) designed for jamming, over three major design-test cycles. Over these three iterations the interface evolved from a very simple 'process control' interface in v1 to a more expressive multi-touch sample manipulation tool in v3. At each stage of the design process, open-ended jam sessions held with local musicians suggested that the potential was there for the interface to support rich jamming experiences. Version 3 of the interface and the associated v3 jam session was the most in-depth of the three phases of the experiment, with the most expressive interface and also the most comprehensive field trial (using a multi-session longitudinal study of jamming musicians rather than the single jam sessions of v1 and v2).

Situating the qualitative results of these experiments within the broader context of third wave HCI, this thesis discusses *affect* in a guise perhaps unfamiliar to readers of mainstream HCI discourse. The jam sessions were characterised by intense sonic atmospheres, and the post-jam interviews reveal a complicated picture of agency in the interaction of the musician and their sound. The thesis also presents a detailed analysis of the quantitative log data, including the results of a Machine Learning (ML) approach to looking for patterns in this data.

Finally, the thesis discusses the implications of the Viscotheque design process for HCI more broadly, including the powerful affective atmospheres which characterise musical interaction and an approach to data analysis which leverages the mathematical sophistication of modern ML techniques while remaining sensitive to the difficulties surrounding the measurement of experience.

Contents

List of Figures

List of Tables

Abbreviations

NB Naive Bayes. 157, 160, 162

NIME New Interfaces for Musical Expression. 19, 43, 44

OSC Open Sound Control. xiv, xv, 49, 50, 53–55, 61, 74–76, 80, 82, 107, 108, 113, 114, 178

PD Participatory Design. 15

RF Random Forests. xvii, xix, xx, 157, 158, 160, 162, 164, 168–170

RMS Root Mean Square. xvi, 49, 148, 156, 157, 188

SLOC Source Lines of Code. 75, 107

SVM Support Vector Machine. 157, 158, 160, 162

TCP Transmission Control Protocol. 53, 55

TEI Tangible and Embedded Interaction. 10

UDP User Datagram Protocol. 53

UX User Experience. 18, 19, 21, 24, 173, 175, 177, 181

v1 Viscotheque version 1. 1

v2 Viscotheque version 2. 1

v3 Viscotheque version 3. 1

VCR Video-Cued Recall. xix, 82, 84, 89, 97, 98, 102, 114

WIMP Windows, Icons, Menus, and Pointing device. 22

WLAN Wireless Local Area Network. xiv, 49, 55, 74, 76

1. Introduction

Open-ended human-computer interactions, such as those in interactive digital art and music, are an increasingly popular area of study in Human-Computer Interaction (HCI). They provide an opportunity to examine playfulness, creativity and expression and challenge conventional HCI notions of quality, evaluation and how to measure success.

Jamming—improvisational group music making—is often held up as the romantic ideal of open-ended creativity. In improvising, jamming musicians spurn pre-planning and give themselves over to the *moment* of the jam, channelling deep and unconscious creative forces as they play. This picture *is* misleadingly poetic—there is far more repetition and structure in jamming than laypeople realise—but it is certainly true that the wild, untamed improvisation of jamming is part of its mystique.

This thesis describes the development of *Viscotheque*, an iPhone-based Digital Musical Instrument (DMI) designed for jamming. This instrument has gone through three major design-test cycles during my PhD programme, which I shall refer to as Viscotheque version 1 (v1), Viscotheque version 2 (v2) and Viscotheque version 3 (v3) (see fig. 10.1). The lessons learnt during this process represent one main contribution of this thesis.

Chapter 2 digresses temporarily from the topic of jamming to discuss third wave HCI. This chapter examines the general concerns and theories currently under debate within HCI discourse regarding open-ended and experientially-oriented human computer interaction.

Chapter 3 then returns to the topic of jamming, primarily as it has been studied from an ethnomusicological and psychological perspective. Focusing on jazz (perhaps the canonical example of improvisational music making), this chapter examines both the low-level and high-level factors which influence the behaviour of jamming musicians and groups. The chapter finishes by discussing the topic of DMIs and their potential as instruments for jamming.

Chapters 4 to 8 describe the evolution of Viscotheque as a DMI for jamming. The system was designed in an iterative, participatory fashion, with each version incorporating both expert musical judgements as well as feedback from musicians in previous studies. The overreaching philosophy of the Viscotheque design process can be summed up in the two phase feedback loop shown in fig. 1.2.

Each version of the system was built and a group (or groups) of musicians were invited to come and jam using the instrument. Few instructions were given to the musician and few constraints imposed—the purpose of these 'field trials' was to see what jamming practises arose organically amongst the musicians. The lessons learnt from these field trials, both from observing the instrument in use and also from interviewing the musicians regarding their experience, were then used to improve the design of the instrument.

Each of these chapters is structured as follows:

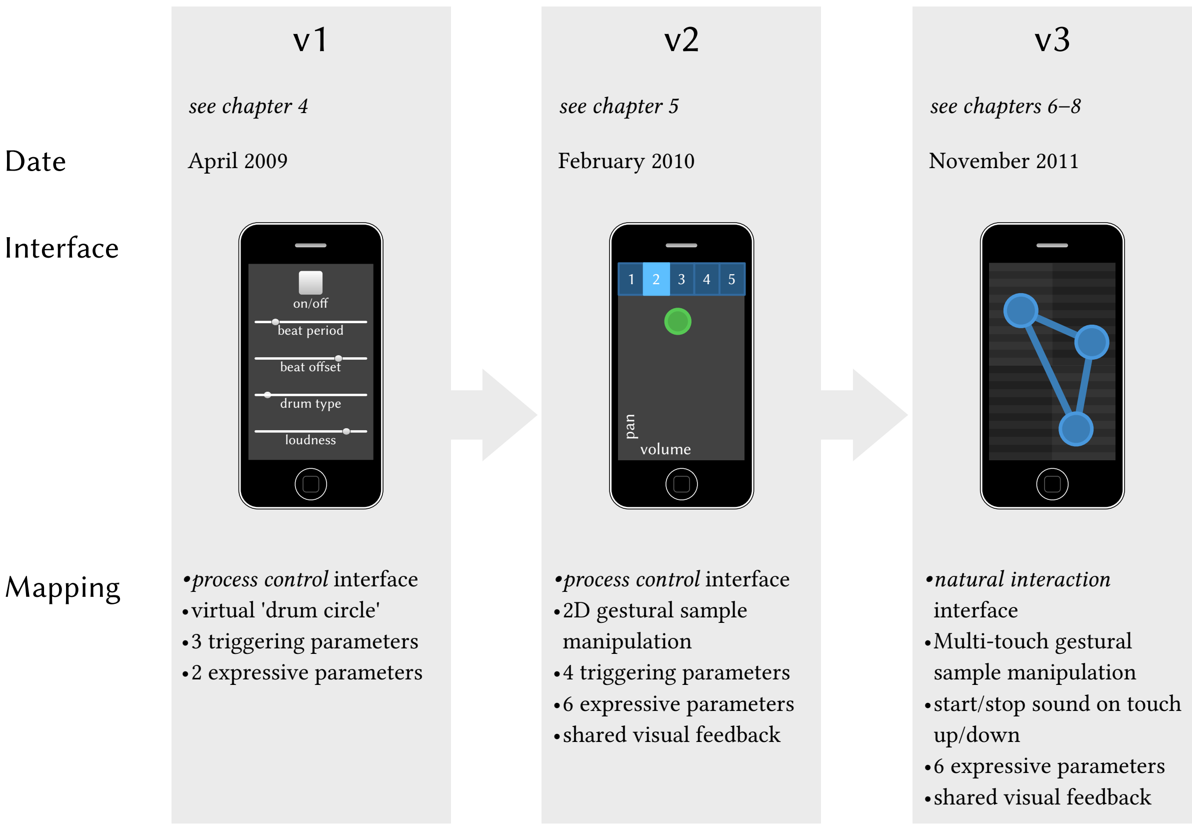


Figure 1.1.: The development of the Viscotheque instrument over the three versions, showing changes to the interface and to the sound mapping.

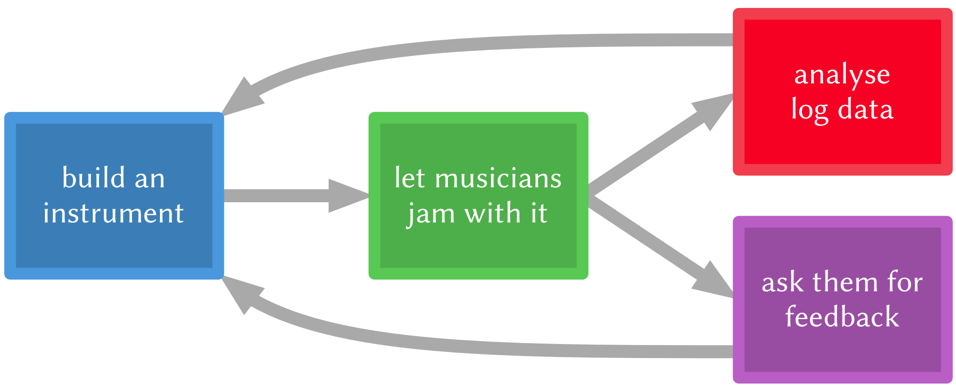


Figure 1.2.: The Viscotheque design process.

Architecture
This section includes technical details about the hardware and software powering Viscotheque and the control flow and musical output of the system. Although the iPhone[1] remained a constant, the physical environment differed between different versions of the system—the setup of the jamming room and the audiovisual feedback provided to the musicians.

Mapping
The mapping between input to the devices sensors (touchscreen, accelerometer, GPS, microphone etc.) and the output sound and visuals produced represents the key space for creativity and innovation in DMI design (Miranda and Wanderley, 2006). For this reason, I shall describe the mapping used in each version of Viscotheque in detail, explaining the rationale behind each design decision.

Jam session structure
Each version of the interface was used in a jam session by musicians familiar with the art of jamming. This section describes the number of musicians that took part in the jam sessions and the experimental protocol used in letting them jam and gathering their feedback.

Data visualisation and analysis
I will then present analysis of the behaviour of the musicians in the jam session as captured by the logged data from the jam. I have deliberately avoided proposing a quantitative measure of the quality of the jamming experience, however there are still interesting insights to be gleaned from looking at the log data.

Musician interviews
The most important aspect of this design cycle was the post-jam interviews with the musicians. These interviews attempted to capture each musician's subjective experience of jamming with the Viscotheque instrument, and to get feedback on the strengths and weaknesses of the instrument as a tool for jamming.

Version 3 of the interface and the associated v3 jam session was the most in-depth of the three phases of the experiment, with the most expressive interface and also the most comprehensive field trial (using a multi-session longitudinal study of jamming musicians rather than the one-off jam session of v1 and v2). For this reason, the findings from the v3 jam sessions are split into two separate chapters. Chapter 6 begins with a discussion of *affect* in a guise perhaps unfamiliar to readers of mainstream HCI discourse before mobilising those ideas in a discussion of the qualitative results from the v3 jam sessions. Chapter 8, in contrast, contains a detailed quantitative analysis of the v3 log data, including the results of a ML approach to looking for patterns in this data.

Finally, in chapter 9 I discuss the implications of the Viscotheque design process for HCI more broadly, including the powerful affective atmospheres which characterise musical interaction and an approach to data analysis which leverages the mathematical sophistication of modern ML techniques while remaining sensitive to the issues raised by the third wave in measuring experience.

[1] Actually, any iOS device (iPhone, iPod Touch, or iPad) can be used as a Viscotheque instrument.

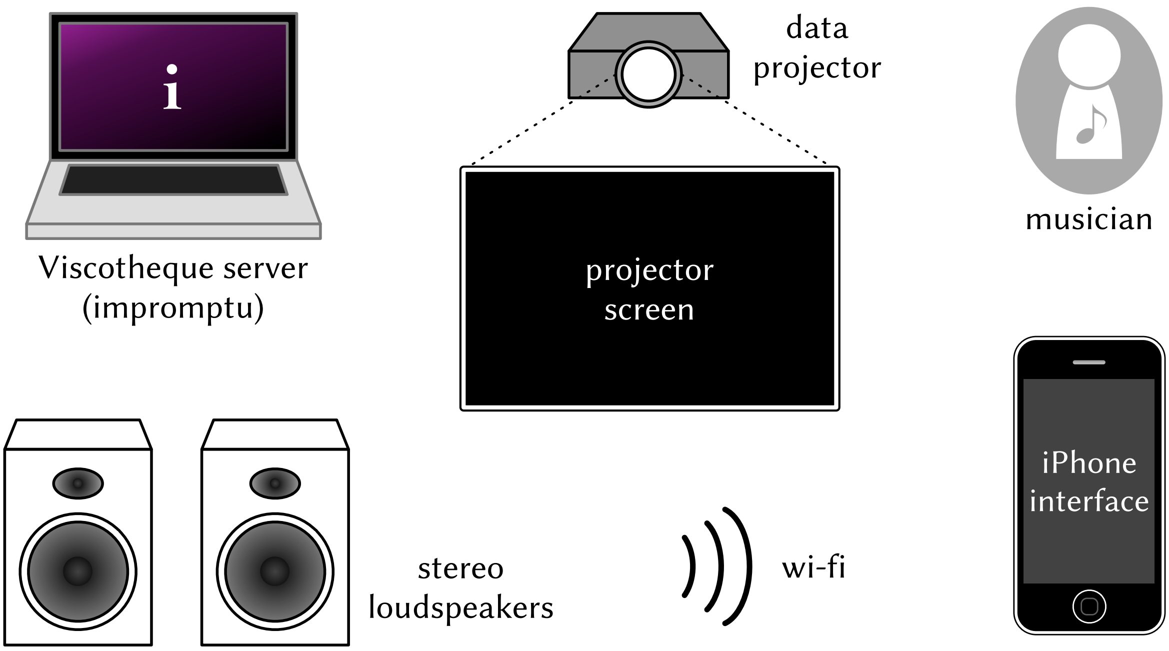


Figure 1.3.: A 'visual glossary' of the symbols used in the figures in this thesis. The system architecture diagrams (figs. 4.1(a), 5.1 and 6.1) use a consistent iconography to represent the various software and hardware components of the system.

2. The road to third-wave HCI

The road is long,
with many a winding turn
That leads us to who
knows where, who knows where?

*(*The Hollies*, He 'Aint Heavy)*

To set the scene, this chapter discusses the 'third wave' (Bødker, 2006) or 'third paradigm' (Harrison et al., 2007) of HCI research, a loose collection of ideas and methods which have risen to prominence in the last decade.[1] I shall present some historical background, showing the road the field of HCI has taken to get to this point, and a summary of some of the key thinkers and ideas which characterise the third wave.

Whether these ideas represent a Kuhnian (1970) paradigm shift, or simply a diversification or strand of theory which can coexist with more traditional HCI methods remains to be seen. While there is no clear consensus, there appear to be two primary dimensions to the third wave agenda. The first is a dissatisfaction with controlled, systematic experimentation and quantitative measures of 'performance', and the second is an inversion of the artefact-experience binary in the teleology of socio-technical systems.

Thirty years ago in the 1980s, personal computing was in its infancy. Having individual access to a computer in the home or at the office was largely restricted to hobbyists, early adopters or employees of large corporations. These computers, even if they didn't require manual assembly (which some did), were unfamiliar and difficult to learn and use. Using a program required juggling the program and operating systems disks, as well as consulting a large (hard copy) instruction manual. Now, it is true that enormous progress was being made at this time. In general the computer as a tool was limited to a few task domains—word processing, spreadsheets and number crunching. It was primarily an instrument of *work*.

At this time, heavily influenced by cognitive psychology, the 'information processing' paradigm was dominant in HCI. Computer use was conceptualised as a goal-directed, action-reaction cognitive loop, with the user perceiving the state of the world, selecting and then performing an action to achieve the desired goal. Frameworks such as Norman's 'gulfs' of execution and evaluation (Norman, 1988) and the Model Human Processor (MHP) of Card et al. (1986) were proposed to model and predict user behaviour. This was the first wave of HCI research.

Although these ideas were foundational for HCI, they presented a picture of an individual human interacting with a single computer. In this picture the 'world' and the cognitive process exist largely in isolation—disconnected except through the specific, structured display

[1] The 'third wave' nomenclature is used in Bardzell and Bardzell (2011); Bertelsen et al. (2007); Fallman (2011); Kiefer and Collins (2008); Rode (2011); Taylor (2011); Wolf et al. (2006); Ylirisku et al. (2009). Harrison et al. (2007) also contains an appendix called "Alternative Names for the Third Paradigm".

of information presented explicitly in the interface. The 'user' came to the interface with specific goals, and the purpose of the interface was to provide the user with input to best choose an action to accomplish that task. Success was well defined—the achievement of the specific goal, a particular state of the world.

Since that time these models, indeed this general paradigm, have come under sustained critique in HCI discourse. In a panel at the CHI 2003 conference called "Post-cognitivist HCI: second-wave theories", shifts in underlying theory were highlighted:

> Historically, the dominant paradigm in HCI, when it appeared as a field in early 80s, was information processing ('cognitivist') psychology. In recent decades, as the focus of research moved beyond information processing to include how the use of technology emerges in social, cultural and organizational contexts, a variety of conceptual frameworks have been proposed as candidate theoretical foundations for 'second-wave' HCI and CSCW. The purpose of this panel is to articulate similarities and differences between some of the leading 'post-cognitivist' theoretical perspectives: language/action, activity theory, and distributed cognition. (Kaptelinin et al., 2003)

Since that time, a *third* wave of HCI theory has been identified:

> In the early 2000s, to find ways of tackling these new challenges (as well as to break with the theories and methodologies of the second wave), HCI became rapidly interested in issues such as meaning, complexity, culture, emotion, lived experiences, engagement, motivation, and experience—HCI's 'third wave'. (Fallman, 2011)

In part, these shifts were brought on by technological advances in the computational devices themselves. Embedded systems, mobile devices, touch interaction; the development of new technologies has provided different interaction contexts which do not fit so neatly into this single-user-at-a-computer worldview. The critique has come from another angle, as well: from a sociological and ethnographic perspective. These voices highlight the need to consider human-computer interaction in all its messy, lived-in-the-real-world complexity. This perspective asserts that even ostensibly simple human-computer interaction contexts are embedded in rich networks of social and cultural phenomena which influence their unfolding.

Historians and philosophers may scoff at this point—HCI has been around for less than half a century, and computers themselves scarcely longer. How can we be up to our *third* wave already? There may be some substance to their incredulity—there is no consensus on exactly what constitutes this third wave, and as such it can be viewed as a fairly arbitrary delineation. However, there is no doubt that computational technologies have rapidly changed the world we live in, beginning with the mainframe, through the personal computer era, and now in the era of mobile and pervasive computing. It is unsurprising that HCI theory has itself undergone twists and turns in an attempt to keep pace with these changes.

2.1. The technology of the third wave

> In young research fields there is often a tendency to be highly opportunity and technology driven and to focus primarily on producing solutions while reflecting less on methodology. (Kjeldskov and Graham, 2003)

It is clear from perusing any electronics catalogue that the 'single user operating an isolated desktop computer' is no longer the only context for human-computer interaction. Technological progress has multiplied the connections between humans and computers (Benyon et al., 2010). Poppe and Rienks (2007) identify four trends in emerging HCI:

1. *New sensing possibilities*: many systems are exploring input devices and modalities beyond the traditional keyboard and mouse.

2. *A shift in initiative*: the user is no longer the instigator of all the action in human-computer interaction. Proactive and provocative actions initiated by the system are becoming more common.

3. *Diversifying physical interfaces*: the physical forms of interfaces are diversifying as well.

4. *Shift in application purpose*: there is a shift away from task-based interaction, towards opportunistic and discretionary use.

Computing devices continue to get smaller, and lighter and more *mobile*. Throughout 40 years of semiconductor research and development, transistor density (which directly impacts upon processing power, efficiency and device size) has grown exponentially, following the predictions of Moore's law (Moore, 1975). Although there are always those willing to predict the end of this era, their worries have thus far proved unfounded.

As a consequence of the increasing portability of computing devices, the physical environments in which human-computer interactions occur have expanded. Ubiquitous computing (Weiser, 1991) and pervasive computing (Dey et al., 2001) describe the era of embedded computing already upon us, with devices measuring, controlling and augmenting our physical environment. Wireless sensor networks for monitoring rainfall and soil quality in agriculture (Hu et al., 2010), smart appliances in the home (Baeg et al., 2007) and cameras which monitor elderly residents and automatically call for help when required (Foroughi et al., 2008) are just some examples of the way in which computing power is being integrated into the environment.

Many of these computing devices are not designed for conscious, intentional interaction. They are passive, simply monitoring the state of the world in various ways. They spring into action when necessary, transmitting information or directly acting upon the world as appropriate. They are aware of (and may be responsive to) human action, but our interaction with them is distinctly different from the wilful interaction of the single-user-at-the-desktop scenario.

This is also true of Tangible and Embedded Interaction (TEI). The computing devices embedded in the environment need not be limited to passively monitoring the state of the world. There are opportunities to connect the tangible world of things with the intangible world of information in new ways in an era of 'tangible bits' (Ishii and Ullmer, 1997). Computationally augmented objects can be used to display information (Brewer et al., 2007), an ambient representation of information for either direct or peripheral stimulation. Augmented objects may also constitute the control surface of a system—to be manipulated in interaction with the system. Examples of this are the Jam-O-Drum (Blaine and Perkis, 2000) and Reactable (Jordà et al., 2007) physical interfaces for music-making. I shall return to the subject of socio-technical environments for music making in section 3.4.

These developments are not confined to the world of academic research. Home brew and DIY electronics are more popular and accessible than ever before, thanks to the availability of development platforms such as Arduino (*Arduino*). The overarching trend is towards a broader range of human-computer connections and a deeper entanglement of the human, the computer and the environment. This is one of the motivating forces behind the third wave shift HCI beyond the individual and beyond the workplace.

The increasing mobility of computing devices is dramatically reshaping our conscious interaction with computers as well. This is nowhere more evident than with the internet-enabled smartphone. As of 2011, 1.7 billion people (68% of internet users, which is also 29% of all mobile phone subscribers) use their mobile phones at least some of the time to access the web or email (Ahonen, 2011). Tablet computers such as the Apple iPad, with 55 million units sold since its introduction in April 2010 (*Apple Reports First Quarter Results* 2012), are replacing desktop computers for many computer users, particularly for email, web surfing and social networking on sites such as Facebook.

Conferences such as MobileHCI provide a showcase of current HCI research in this area (Kjeldskov and Graham, 2003). The research agendas of this community are diverse; from information presentation issues (Church and Smyth, 2009; Gostner et al., 2008) to the use of mobile devices for live group remixing of multimedia (Cao et al., 2007; Vihavainen et al., 2011). The array of input and output modalities provided by modern smartphones continues to grow, including multi-touch screens, location, orientation and movement sensors, cameras, microphones, loudspeakers and more (Ballagas et al., 2008).

The mobile phone remains an artefact for communication as well. Indeed, such devices are changing the way we communicate, altering geographies of information and notions of distance and connectivity (Sheller and Urry, 2006). The near-ubiquity of high-speed wireless network coverage in metropolitan areas means that these computing devices are always connected to the internet and to one another. If the portability of these devices has moved computing off the desktop, then their constant connectedness has moved computing beyond the individual as well (see fig. 2.1). Even on the 'boring' desktop, technological advances have spawned new contexts for HCI. Trackpads and multi-touch trackpads are now available in mainstream commercial and open source operating systems. Multi-modal interaction via face and voice tracking (Jaimes and Sebe, 2007) and even bio-signals such as EEG (Wolpaw et al., 2002) offer additional communication channels in interface design.

Commenting on how these technological changes have changed the field of HCI, Rogers writes

> One of the main reasons for the dramatic change in direction in HCI is as a reaction to the explosion of new challenges confronting it. The arrival and rapid pace of technological

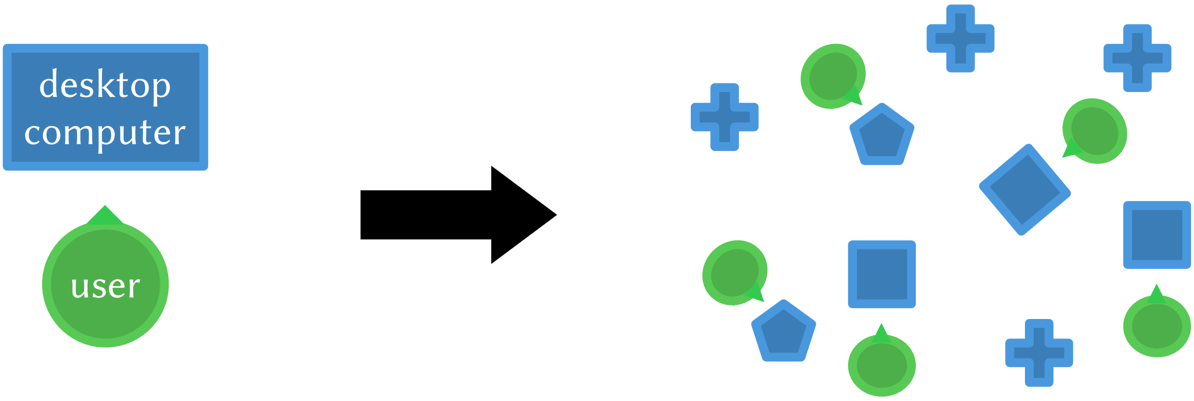


Figure 2.1.: Over time, the focus of HCI research has moved away from the single-user-at-a-desktop paradigm to multiple users, multiple (non-desktop) computing devices, home and leisure environments, etc.

> developments in the last few years (e.g. the internet, wireless technologies, hand-held computers, wearables, pervasive technologies, tracking devices) has led to an escalation of new opportunities for augmenting, extending and supporting user experiences, interactions and communications. These include designing experiences for all manner of people (and not just users) in all manner of settings doing all manner of things. (Rogers, 2004)

In simplistic terms, the forces of mobility and connectivity are shaping an increasingly interconnected human-computer landscape. This is the *technological* context in which the third wave is situated, but there have also been other reasons for the proposal of new perspectives in HCI theory. The second and third waves of HCI have also been driven by an ethnographic and anthropological critique.

2.2. The (cyber)anthropology of the third wave

Historically, there has been a steady flow of ideas from other disciplines into HCI theory. There are several reasons for this. HCI, being relatively young as an academic discipline, has a comparatively small body of existing literature and thought. This also means that there is low-hanging intellectual fruit to be plucked. Also, as computing technologies are incorporated into human behaviour, then any field which is concerned with human behaviour on any scale (sociology, anthropology, psychology, etc.) can bring its ideas to bear on the human-computer interaction. Finally (and cynically) it may be the case that computer science departments, flush with grant money and industry partnerships, offer a safe haven from the storms of budgetary austerity which plague other areas of academia. Whatever the reason, HCI is truly a melting pot—a hectic bazaar where an eager researcher is sure to be able to find a theory or framework which suits their purposes.

In the first wave the conception of the human was strongly rationalist, with a focus on mechanistic models of cognition: where a human agent sets a goal, then plans and executes a series of actions to achieve those goals. Detailed models were formulated to predict the amount of time required to perform certain well defined tasks (Card et al., 1986). There was a strong emphasis on empirically derived numerical relationships, such as Fitts' law for target acquisition with pointing devices. (Carroll and Rosson, 2007). Subsequent sociological critique has raised questions such as what is a human; what is a computer? This is not just a debate about methodology, it is a debate about the very essence of human-computer interaction. These critiques are a challenge to consider the importance of connections and avenues of influence which are ignored in the cognitivist paradigm. This section presents a brief overview of this expanding horizon of human-computer interconnections, beginning with the second wave focus on context and ending with the third wave's emphasis on experience and examining interaction in the wild.

Understanding Computers and Cognition: A New Foundation for Design, by Winograd and Flores (1986) was a seminal critique of the prevailing rationalist-cognitivist atmosphere in HCI. Their primary criticism was directed towards Artificial Intelligence (AI) approaches which treated cognition as the manipulation of symbolic representations of the world. Computers, they argued, do not (and cannot) have access to the shared background of understanding which makes human language and communication *robust*. The shared history and culture which allows human communication to go 'off script' and beyond the anticipated is not amenable to representation in any form that computing devices are capable of.

This was a strong critique of the 'expert systems' approach to AI. Their claim was that human behaviour and intelligence is dependent on factors which cannot be encoded in a way the computer can understand. This is not to say that computers are not a crucial and enabling part of the human-computer interaction, merely that intelligence cannot be reduced to a system of explicit '**if** *context* **then** *action*' contingencies.

Suchman's (1987) *situated action* has also had a significant impact on HCI discourse. While working at Xerox PARC, Suchman observed pairs of users interacting with an expert help system designed for a photocopier which was notoriously difficult to use. Adopting an ethnographic approach, she observed users *in situ* and performed conversation analysis on video recordings of these interactions.

Suchman's primary criticism and warning was against considering the user's behaviour only in the abstract. She stressed the *situatedness* of action, its dependence on the social and material circumstances of the actor. Unfolding patterns of behaviour in human-computer interaction are not governed by an abstract model of rational user behaviour or a linear perception-cognition-action model. They are embedded in real-world situations, against an unarticulated background of experiences and situations.

> Activity is fundamentally improvised; contingency is the central phenomenon. People conduct their activity by continually re-deciding what to do. (Agre, 1997, chapter 1)

Like Winograd and Flores (1986), Suchman also framed human-computer interaction in terms of a communication asymmetry. Humans possess a wide range of linguistic, nonverbal and inferential resources with which make sense of and act in the world. Computers, on the other hand, have a limited array of sensors and actuators with which to communicate. The problem of design, then, is to find ways to mitigate this communication asymmetry (L. Suchman, 2006, p179).

Winograd and Flores and L. Suchman were key figures in the second wave of HCI theory (Fallman, 2011). Suchman's ethnomethodological approach (following Garfinkel (1967)) of observing the way that people make sense of their world as they go about their way in it, has had a lasting influence in HCI (Brodersen and Kristensen, 2004). The study of human-computer interaction 'in the wild' as it occurs in the real world (rather than in the laboratory) is a repeated theme in third wave rhetoric also.

There are critics of this 'context-heavy' approach. Nardi (1995) warns that a hard line insistence on the importance of the *specifics* of a given human-computer interaction context leads to an account so dense with detail as to be useless for uncovering more general principles. To deal with this difficulty, qualitative analysis techniques such as Grounded Theory (Strauss and Corbin, 1990) are often used to derive normative laws from qualitative data (Furniss et al., 2011).

The primary contribution of this second wave ethnomethodological critique is that the social and material circumstances of human-computer interaction *matter*. They often subvert or derail any pre-formed plans and they are performed *in the moment*. A proper understanding of the human-computer interaction must address not only the human and computer in the abstract, but the concrete situation and practices in which they are embedded.

Moving on to the third wave, technology is pushing beyond well-defined individual and group contexts and into the environment in the form of tangible computing. This is but one example of a potential third wave context, but it highlights the ill-defined nature of interaction, in temporal and spatial and intentional dimensions. Dourish (2001b), an influential voice in third wave HCI, views the twin strands of tangible computing and the focus on the social and cultural context of computing as having a common grounding in the concept of embodiment.

> By embodiment, in this context, I mean not simply physical presence, although that is certainly one relevant facet. More generally, however, by embodiment I mean a presence and participation in the world, real-time and real-space, here and now. Embodiment denotes a participative status, the presence and occurrentness of a phenomenon in the world. So, physical objects are certainly embodied, but so are conversations and actions. They are things that unfold in the world, and whose fundamental nature depends on their properties as features of the world rather than as abstractions. (Dourish, 2001a)

Drawing on the phenomenology of Husserl, Heidegger, and Schultz, Dourish recognises the 'being-in-the-world-ness' of *all* human-computer interaction. To fully understand the human-computer assemblage, he argues, one must see it as fundamentally embodied and embedded in the world—and not just a world of corporeal bodies but of conversations and ideas as well.

Dourish explicates two main implications of this position: embodiment is about establishing meaning, and meaning arises in the course of action. This constructivist relation between meaning and action is a common theme in third wave HCI. Humans in interactive systems are engaged with the world, and sense making is a process, not a destination. Meaning is not deterministic, encoded into artefacts by designers to be 'discovered' by users. Meaning is created in encounters—it is *per*formed, not pre-formed.

Like Suchman, Dourish is pushing for the consideration of *context* in human-computer interaction. The word context is often the site of misunderstanding among designers of interactive systems. Like Winograd and Flores (1986) before him, he argues that context is not merely an attribute of the environment, something else which can be measured and represented by a sufficiently aware system[2] (Dourish, 2004). Context is enacted, it is in the interplay of beliefs, experiences and dispositions that the different parties use as they interact. It cannot be captured or put in a bottle. The distinction between *space* (a material reality) and *place* (whose meaning is socially enacted) is one example of this (Williams et al., 2005). Spatial location can be measured with GPS, but no electronic device can observe the complex history and social influences which make a place meaningful.

As Fallman (2011) notes, HCI has increasingly wrestled with the notion of 'the good' as it has undergone these twists and turns. Although questions of the good may seem like yet new challenges, in truth they have always been implicit in the HCI project (Gilmore et al., 2008). A simplistic articulation of the aim of the field could be given as 'to make interactions between humans and computers *better*'. The battleground, as always, is the definition of the word 'better'.

Participatory Design (PD) acknowledges the capacity of technology to transform labour relations (Schuler and Namioka, 1993). PD is committed to giving workers 'on the ground' control over the computational tools (the means of production) used in the workplace. This overtly political stance is congruous with the ideological atmosphere of the Scandinavian nations where it originated. The driving force here is an ethical one—that interaction design is not neutral in regard to the distribution of labour, and that these connections must be taken into account.

This is an example of another commitment of the third wave—the consideration of *other* perspectives on technology use. That is, the consideration of the use (and benefits) of technology outside of the 'middle-aged, white, educated, neoliberal male' hegemony. Irani et al. (2010) propose the lens of 'postcolonial computing' for understanding design in encounters with "other" cultures, including issues of uneven economic relations and development. This work—so called ICT4D—involves investigating how computing devices are changing traditional cultural practices and raising living standards in the developing world.

HCI has also seen (along with the rest of the world, particularly the developed nations) an increase in discussion of environmental sustainability in recent years (Mankoff et al., 2007). DiSalvo et al. (2010) classifies this body of literature into several different approaches to the problem of sustainability "providing key topics for future research and reflective discussion in sustainable HCI" (ibid.). Indeed, and at the risk of invoking an alarmist "won't somebody *please* think of the children" pathos, it is clear that technology has considerable potential to address the environmental challenges facing our planet. Taylor is critical of the way that third wave HCI has rushed to consider these other (particularly non-western) cultures, of these other cultures, 'exoticising' them in the quest to find something different. This is again part of the third wave emphasis on the importance of cultural factors on the way that human-computer interactions unfold.

> I can't help but wonder whether a number of us in HCI are caught up in an exercise

[2] This 'sufficiently aware system' may well be HCI's 'sufficiently smart compiler'.

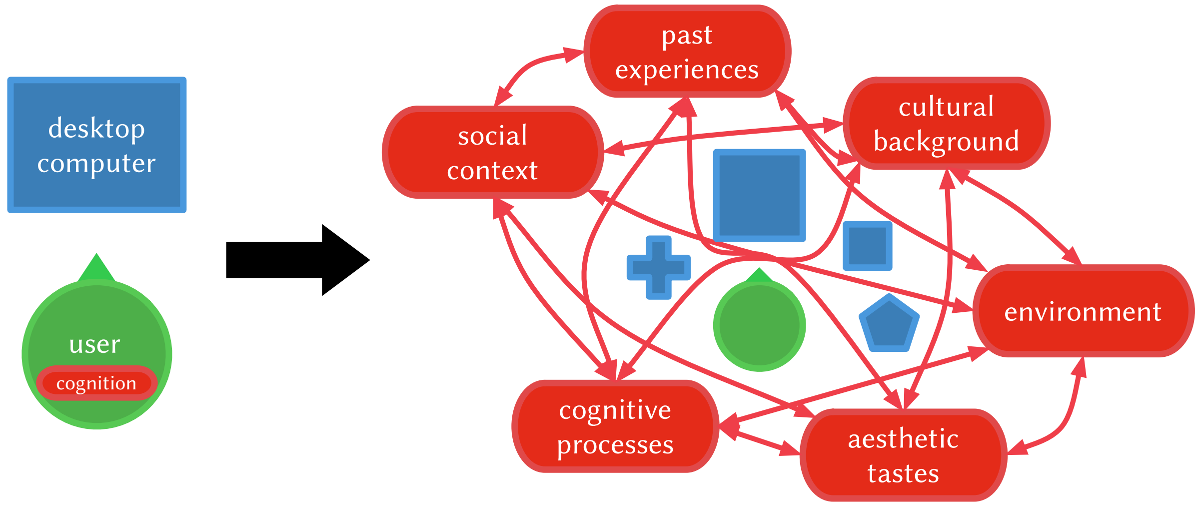


Figure 2.2.: As well as the move towards group computing and a variety of computing devices, HCI's holistic concern with the user beyond a purely cognitive understanding has included a consideration of various social, environmental and cultural aspects of any human-computer interaction (the concepts shown in the figure are not an exhaustive list.

> of repeatedly looking further afield so that we can report back that things are different out there, that people's ways of knowing and practices are culturally situated, and, furthermore, that their activities are interconnected in complex ways. These are, of course, important points, but as a field that strives to produce new and hopefully provocative perspectives on human-computer interactions it seems we should be aiming to introduce something else besides these familiar ways of seeing. (Taylor, 2011)

The purpose of this section is not to promote any particular ethical position, but merely to point out that third wave HCI discourse is increasingly concerned with the 'big questions' (see fig. 2.2. The key shifts in perspective on what *matters* in human-computer interaction have tended towards holism over reductionism, towards interdependence over independence, and towards connectionism over isolationism.

2.3. *Experience*: the catch-cry of the third wave

One feature of the third wave is the expansion of computing beyond the workplace to include a "focus on the home, on leisure" (Bødker, 2006). This is in contrast to the first and second waves of HCI theory, which were concerned with individual and group work contexts respectively (Fallman, 2011). As discussed in section 2.1, there are an increasingly diverse range of human-computer environments being designed to facilitate exploration, creativity and play (Costello and Edmonds, 2009). Indeed, this dissertation presents just such a system—the group improvisational music making environment *Viscotheque*.

A by-product of the increased diversity in computing contexts is the increasing *open-endedness* of human-computer interaction. Some tasks involving a human and a computer (and potentially many humans and computers) have as their goal the production of an artefact, for example transcribing recorded speech into a word document or entering data into a database. Other tasks involve achieving a certain objective (a certain state of the world), such as scheduling a group meeting without any schedule conflicts or paying a bill online. In these task scenarios the success (or failure) of the humans and computers involved is relatively straightforward to measure. How *good* is the transcript? How efficiently was it typed and formatted, and how many mistakes are present? Similarly, was the bill payed in full and on time? Was there sufficient money in the account? The answers to some of these questions require more interpretation than others, but in general they are meaningfully measurable either as a yes/no answer or a simple (perhaps 1-dimensional) result. Especially in the case of an interaction context where the primary goal is the production of a digital artefact, the success of the activity or and the effectiveness of the interface can be measured by the quality of the artefact produced.

How does one evaluate the success of an interaction, or compare different environments for computer-supported creativity? In wrestling with this question, 'experience' has become a hot topic in HCI over the last decade: 'it's all about the experience' was even the tag-line of the 2012 CHI conference. This trend has been framed as a shift in focus from *usability* (Sauro and Lewis, 2009) to *user experience* (Hassenzahl and Tractinsky, 2005). In this shift, the notion of usability (Nielsen, 1993), which enjoyed a position of primary importance in the cognitivist design literature, is but *one* of the goals of the designer. Other aspects of the human experience, such as pleasure, joy, surprise or frustration are also important aspects of human-computer interaction, and may even be more important than usability in some situations. "Emotions and experiences are keywords in the third wave" (Bødker, 2006). Even Norman, one of the champions of usability engineering, has more recently insisted on the emotional dimension of product use (Norman, 2005). This emphasis on fun and enjoyment is sometimes portrayed as an antidote to 'stodgy' traditional HCI (Monk et al., 2002).

Further dispelling the perception of stodginess, interactive art is an increasingly common setting for human-computer interaction (Morrison et al., 2011), and "computing technology has a vast potential to support interdisciplinary creative collaboration." (Mamykina et al., 2002) In this context interaction is wilful, yet exploratory and improvisational. Indeed, finding a balance between guiding the interaction and letting new and unexpected behaviours occur is one of the key challenges in designing interactive art environments (Costello and Edmonds, 2009).

With computer games, too, the objective of the game is hard to define. While some invariably play for high scores, others suggest that bragging rights (Su, 2010) or a feeling of immersion (Jennett et al., 2008) may be the ultimate goal of the gamer in this human-computer interaction context.

Evaluating the worth of any artistic endeavour is usually performed through expert critique, and the 'test of time' is often the ultimate standard to which art may be held (Edmonds and Candy, 2002). As interactive artworks which include computing technology have been studied in HCI, and these have played a large part in driving the re-evaluation of more traditional evaluation methods in the third wave. These burgeoning creative and improvisational human computer interaction contexts have been given many labels, such as:

- Digital Live Art (DLA) (Sheridan et al., 2007)
- Digital Musical Instrument (DMI) (O'Modhrain, 2011)
- Computer Supported Collaborative Play (CSCP) (Twidale et al., 2005)
- Computer Supported Creativity (CSC) (Knörig, 2007)
- New Interfaces for Musical Expression (NIME) (Dobrian and Koppelman, 2006)
- Computer Mediated creativity (CMC) (Bertelsen et al., 2007)

The use of critique (rather than evaluation) has been suggested as a more general answer to the difficulty of incorporating the cultural and experiential context into the design process. Bardzell and Bardzell have proposed an analogous 'interaction criticism'. Drawing on 20th century critical theory, they prescribe

> interpretive analysis that explicates relationships among elements of an interface and the meanings, affects, moods, and intuitions they produce in the people that interact with them (Bardzell and Bardzell, 2008).

Interaction criticism proposes four loci of analysis: the designer, artefact, user, and social context. These elements are all deeply interconnected, the aim of interaction criticism is not to claim any independence between them. Rather, they claim to provide a common basis and vocabulary for examining interactive digital environments. The appropriation of ideas from literary (Blythe, 2004), critical and cultural theory (Blythe et al., 2010; Satchell, 2008) is a feature of third wave approaches.

There is a growing body of work concerning User Experience (UX), in which the I of 'interface' in User Interface (UI) is replaced by the X of 'eXperience' (see (Forlizzi and Battarbee, 2004), for example). This change acknowledges the fact that in human-machine interactions it is not just the *interface* that is important, it is the *whole experience.* In this literature there is a diversity of views about what the term 'experience' means (Law et al., 2009) and how to measure it. According to Law (2011), the work on experience can be broadly divided into two camps: the 'reductionists' and the 'holists'. The reductionists, coming from a cognitive psychology perspective, see experience as a *quantity*; a vector in an ill-defined, high-dimensional phase space. Their research problem, then, is to estimate this vector (and perhaps its trajectory over time) in a given human-computer interaction context. Often this is done using questionnaires (Finstad, 2010; Hassenzahl, Diefenbach, et al., 2010; Schaik and Ling, 2008; Väänänen-Vainio-Mattila, 2009) or post-interaction interviews (Burmester et al., 2010). In this literature, a positive user experience is fundamentally related to need satisfaction (Hassenzahl, 2008), where a taxonomy of human needs is provided by the psychological literature (for example Maslow, 1954; Sheldon et al., 2001).

A different vision of *experience* and human-computer interaction is provided by McCarthy and Wright (2004) in *Technology as Experience.* Drawing on the work of pragmatist philosopher John Dewey (in particular *Art as Experience*, 1934) and literary critic Mikhail Bakhtin, McCarthy and Wright present a more holistic picture of experience and technology, which has the same interactionist emphasis as some of the second wave ideas (e.g. Suchman's situated action), but with a greater emphasis on the emotional-volitional aspect of this interaction.

> For us, felt experience points to the emotional and sensual quality of experience. Our first proposition is that these qualities should be central to our understanding of [the] experience of living with technology. (McCarthy and Wright, 2004, p13)

In this perspective, human-computer interactions happen against a backdrop of previous experiences, desires and passions.

McCarthy and Wright suggest that all experience is created in dialogue between 'centres of value', which may be humans, groups, ideas or devices. Emotions, for instance, do not exist in isolation—they have a cause, and they also have knock-on effects. The authors use the language of 'threads' (rather than components) to talk about experience in an effort to emphasise the interweaving of different factors in the construction of meaning. The four proposed threads of experience are the sensual, emotional, compositional and spatiotemporal. These threads are not objective facts, they are ideas to help the theorist think more clearly about the experience of interacting with technology (ibid., p79).

Apart from the emphasis on emotional-volitional dimension human-computer interaction, the other key aspect of this work is the focus on *process*. Interactions do not reach completion, they are always subject to re-interpretation and re-configuration.

> Thus, action retains its eventness by always being open to the future, and its meanings are indeterminate. In such an open, free characterisation of human action, action is always potential and always becoming, constituted dialogically in responsive relations. (ibid., p71)

Even seemingly well-defined tasks, such as sending an email or making a phone call, have an influence beyond the immediate temporal and material circumstances.

The 'threads of experience' framework has also been used as explanatory framework by other authors (Leong et al., 2010; O'Brien and Toms, 2008; Swallow et al., 2011; Taylor et al., 2011; Wright et al., 2008). The framework is firmly anchored in the richly descriptive qualitative tradition, with detailed accounts, small sample sizes, and few quantitative results.

The holistic approach to experience asserts that good experiences do not follow deterministically from the system design and initial conditions of the interaction.

> Enchanting experiences may be designed only by approaching enchantment obliquely: not by engineering it in, but by providing opportunities where it may emerge. (Sengers, Boehner, et al., 2008)

The mystique and ephemerality of experience contributes to the perception that design is the 'black art' of HCI (Wolf et al., 2006).

It is the insistence on the difficulty of 'measuring' felt life which puts this tradition at odds with the reductionist interpretation of UX (Law, 2011). Even though both the reductionists and the holists identify experience as the focus of their work, the former are much more closely aligned to traditional HCI research practices, while the latter, with their literary influences and emphasis on the complexity of interaction, are firmly in the third wave camp. This has been the source of some debate within the UX community itself:

> We need models, theories and representations to capture and communicate ideas about designing for and evaluating UX, but they are neither the reality nor are they unbiased. Being aware of this is crucial. Irrespective of whether strict formal measurement paradigms are brought to bear on traditional HCI phenomena like usability or emerging ones like user experience, it is the *persuasiveness* of empirical evidence that is ultimately the test of its worth. Furthermore, one may argue that basically everything can be measured, but some things may be more "measurable" than the others; how to estimate the threshold of measurability remains unclear. Above all, measures need to be meaningful, valid and useful. (ibid.)

It is unfair to label all of this third wave research as quantification-averse (or numerically challenged), but a broad generalisation which holds in most cases is that the reductionists assert that experience can be assigned a number, while the holists assert that it cannot.

This an example of the broader point that in third wave thinking the influence of social and cultural factors on a given human-computer interaction context are not problems to be mitigated but connections to be celebrated, because the goal of the system designer is no longer usability or lower task completion times or anything so mundane.

> And it is to celebrate the interaction of social, cultural, and environmental factors for constituting us and making life meaningful. If interconnectedness is a fact, then interdependence is a goal. (Light, 2011)

The emphasis on connectedness is not relegated to the connections between users in group interaction, there is a commitment to considering the relationship between the designer and the user.

> By creating contexts for design that arise through being attentive to experiences that emerge in the here and now, we have aimed to avoid disenfranchising those we design with by avoiding an approach that relies on satisfying predetermined outcomes. In this way, the designer must view interaction from a holistic rather than specific perspective and seek to understand user's everyday practices and the prior knowledge they bring to a situation…The unexpected and irrational becomes inspiration for creativity. (Frauenberger and Good, 2010)

This is not a one-way relationship, either. The designer must be influenced and inspired by the life their artefact takes on in its use because of the contention that meaning is constructed in situ.

There is a third wave emphasis on studying interaction 'in the wild' (Brown et al., 2011; Hinrichs and Carpendale, 2011; Linden et al., 2011). Brown et al. gives a helpful overview of the challenges and opportunities presented by field trial methods, noting that

> trials and naturalistic deployments of systems have become a core method for investigating user interactions with systems (Brown et al., 2011).

This mention of a naturalistic view of systems in use gets to the heart of the appeal of the in the wild field trial. The unstated (but implied) point is that laboratory studies are somehow artificial and unrepresentative of the way that these complex human-computer interactions take place with real users. Motivating the push towards doing research in the wild is the idea that users actions are situated in and response to their specific socio-material environment, and that when the user is outside of this environment (such as in a laboratory HCI study) their actions will then necessarily then be different. The users may well surprise the designers in their interaction with the product in the wild, leading to new insights and better designs (Pousman et al., 2008).

One interesting approach to studying users in the wild is the concept of cultural probes (Gaver et al., 1999). Gaver and his colleagues left documents and trinkets behind as they visited elderly members of the community, with the idea that the elderly folks would ruminate on them as 'stimulus material' and send them back to the researchers after a period of time.

> The cultural probes—these packages of maps, postcards, and other materials—were designed to provoke inspirational responses from elderly people in diverse communities. (ibid.)

Now this was not a *study* of human-computer interaction, the purpose of the exercises was to 'probe' the users in their natural environment, and ascertain what interested them and what stimulated them with a view to using this data as an inspiration for design. This approach is a long way from the laboratory approach of watching users through a one-way mirror, and is an example of the third wave commitment to understanding the social and cultural context of interaction.

It is important to note that these third wave ideas have not replaced more traditional HCI research. Fitts' law is alive and well, and a great deal of interesting research continues into more 'mundane' human-computer interaction contexts and systems. In HCI conferences papers about the philosophical implications of quantifying experience and the performative effect of gender on mobile phone use in sub-Saharan Africa coexist with papers about improving the design of Windows, Icons, Menus, and Pointing device (WIMP) interfaces. So the language of a third 'wave' is only appropriate if we remember the superposition principle—that multiple waves can coexist in space and time, and the overall field is the result of the constructive and destructive patterns of interference between these wave components.

The third wave has detractors within the HCI community. There are still many who hold that the third wave goes too far in its resistance to laboratory trials and quantifiable outcomes. They argue that while a hard-line commitment to the user as a disembodied cognitive process is unhelpful, and while cultural and environmental factors are important, helpful, there is still great benefit in the traditional 'hypothesis→controlled experiment→statistical test' paradigm. As Rogers notes:

> A problem with allowing a field to expand in this eclectic way is that it can easily get out of control. No-one really knows what its purpose is anymore or indeed what criteria to use to assess its contribution and value to knowledge and practice. For example, of all the many new approaches, ideas, methods and goals that are now being proposed how do we know which are acceptable, reliable, useful and generalisable? Moreover, how do researchers and designers, alike, know which of the many tools and techniques to use when doing design and research? What do they use to help make such judgements? (Rogers, 2004, p88)

The core of this critique is one of *utility*—these descriptions of experience as a complex fabric woven of many threads are interesting and very poetic, but how do we judge which of them are useful and generalisable? How do each of these new ideas fit into the deeper, overarching theoretical background of the field? Is there a need for this grounding? Some argue that there is not:

> Because of its emphasis on multiple perspectives, the third paradigm does not espouse a single, correct set of methods or approaches to answer these questions. Instead, we see a variety of approaches that are embedded in a similar epistemological substrate. This substrate is analogous to a biological matrix, a compatible environment that supports the emergence of a heterogeneous variety of specific structures and connects them to one another. (Harrison et al., 2007, p8)

The key point is that the traditional HCI vs third wave debate is not *just* a problem of evaluation. The goalposts have genuinely moved, such that the goal—the *telos*—of human-computer interaction has changed (Springett, 2009). In a word processor, the ultimate goal of the user is the production of a high-quality document. The contribution of HCI theory is to make this task as pleasant an experience as possible. However, in an improvisational computer-music environment, for instance, the goal of the participant(s) is to have an experience: of flow, connection, or 'groove'. The musical output of the system is merely a means to that end. In these two different contexts the role of the *created artefact* and the *experience of making it* are reversed. The highest goal of an interface or and interaction context is no longer to be functional, it must be a joy to use. These two outcomes are not independent (indeed they are strongly correlated in many cases) but the fundamental shift is in their prioritisation: traditionally usability and task performance were the ultimate goal, with experience being used as a 'hygiene factor' (Hassenzahl, Diefenbach, et al., 2010), whereas the third wave assertion is this binary has been inverted. UX is the ultimate goal, and usability is a proximate one. There are simply a wide range of views on what that means and how this shift will affect the systems that we design and study.

2.4. Chapter summary

To summarise, the third wave of HCI theory and practice has been born out of a dissatisfaction with traditional individualistic, work-oriented and cognition-focused perspectives on human-computer interaction. This has been driven partially by developments in the technology (the computer side of the HCI equation), but also by discourse about the nature of the human, which came initially from other research disciplines but have developed into a robust debate within the HCI community.

While there is agreement that hard to define qualities such as user experience are of prime importance in the design of interactive human-computer systems, the debate continues about the best way to represent and analyse these qualities. I shall return to this discussion in chapter 9.

3. The *essence* of jamming

> We're jamming
> I want to jam it with you.
> We're jamming, jamming
> and I hope you like jamming too.
>
> (Bob Marley, *Jamming*)

Picture this scenario: a group of teenagers get together in a garage after school. Each pulls out an instrument—guitar, bass, drums, keyboard—and turns their amplifier up to 11. Without any firm idea about what they'll play, the bassist starts playing, improvising a bassline. The drummer picks up on the rhythm, and the keyboard player and guitarist join in as well. Each musician adds their contribution, and in fitting in with each other new ideas are planted; new sonic spaces explored.

The jam session lasts for four hours. As they emerge from the garage, ears ringing, there is no recording of the music they made, no audience applause, and no paycheck to collect. The reason they get together, the reason they will come back to do it all again the next week is that it feels amazing. *Jamming* takes place in various guises; with variations in the age and skill level of the musicians, the instruments they play, the style or genre of the music, and their physical environment.

This description is perhaps too romanticised; it doesn't mention the moments in the jam session which sound terrible because the teenagers are still learning their instruments, or the frustration of having old hand-me-down equipment fail, or any of the other factors which might conspire to ruin the jam. What I hope the description does is paint a picture of what can happen when musicians get together to jam. Having in the previous chapter discussed in broad terms the debates surrounding third wave HCI, I would like in this chapter to zoom in on the specific activity of jamming.

The *New Oxford English Dictionary* (2010) contains this definition:

> **jamming** *(v. informal)*: improvising with other musicians, especially in jazz or blues.

Jamming has been studied from multiple different angles, including ethnomusicology, psychology, and HCI. In attempting to come up with a watertight definition of the term 'jamming' I am reminded of the famous quote:

> I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it... (Justice Stewart, Jacobellis v. Ohio 1964).

As a working definition, in this chapter I shall use the verb 'jamming' and noun 'jam' or 'jam session' to refer to the practice of improvisational group music-making (see fig. 3.1). Deconstructing this definition (in reverse order), jamming has these characteristics:

Figure 3.1.: Concept Venn diagram for jamming. It is possible to improvise on one's own, for instance, and there are also many non-improvisational group musical practices, but these three attributes are at the heart of what it is to jam.

Music making

Jamming is a form of music making,[1] an activity involving a *musician* playing an *instrument* as a form of 'aural art' (Hamilton, 2007). Defining 'music' is even more difficult than defining jamming, but in general music making is the production of sound(s) as a creative process, in accordance with some form of aesthetic principles.

Group

Jamming is a group activity—involving multiple musicians and instruments. The instruments need not be the same, and indeed often the different group members will fulfil different roles (rhythmic, melodic, or harmonic) based on the strengths and conventions of their particular instrument.

The primary implication of jamming being a group activity is that each individual musician is no longer in complete control—no longer only acting, but *inter*acting with the other members of the group.

Generally these groups are small in size, perhaps two to five musicians, although larger groups are possible. The jamming group come together (in time and space) to jam, and these encounters are called *jam sessions*. The music making happens in a real-time tight feedback loop, where every musician can hear and be heard by every other musician in the jam.

[1] Also, some non-musical activities which revolve around improvisational creativity (such as improvisational theatre) sometimes refer to what they do as jamming (Sawyer, 2003).

Improvisational
In jamming, the music played is composed on-the-spot or in-the-moment, rather than being pre-composed. This is not to say that the music played is random or unplanned, but that the decisions on what specific note to play or sound to make are made *in* the jam session, rather than beforehand. It is this improvisational aspect which gives rise to the open-endedness which is characteristic of jamming.

Understanding jamming as improvisational group music making is an answer to the question '*what* is jamming'. But more interesting are the questions of *why* and *how*. Why do musicians spend hours practising to participate in these jam sessions? How is it that a group of strangers can come together and with minimal verbal negotiation launch into a coordinated exchange of music making, collaborating to make a beautiful sound? What is the *essence* of jamming?

To answer these questions, it is worth looking at the way that jamming has been studied in the academic literature. The canonical example of an improvising group in music is the jazz ensemble (MacDonald and Wilson, 2006). From a simple trio all the way up to a big band ensemble, improvisation is an integral part of what it is to play jazz. For this reason, the majority of the literature on jamming comes from this perspective and examines this tradition. It is helpful to see in this literature key facets or animating principles of jamming: *conversation*, *agitation* and *differentiation*. In this chapter I will use these headings to survey the literature on jamming. Then, in the final section of this chapter, I will examine the effect and influence of digital technologies on the cultural practice of jamming.

Some of the material in this chapter is taken from the book chapter B. Swift (2012b). "Chasing a Feeling: Experience in Computer Supported Jamming". In: *Music and Human-Computer Interaction*. Springer.

3.1. The conversation of jamming

Studies have shown that musicians have their own vocabulary for talking about what they do when they jam together, and this vocabulary can help us to understand the process of jamming as understood and experienced by its practitioners. In Monson's (1997) interviews with professional jazz musicians, the metaphor of dialogue or conversation was used to describe the act of improvising together. In these interviews the high points in jamming were occasions where musicians felt like they were 'saying something'; expressing something meaningful through their playing. The conversation metaphor connotes a sharing of ideas, a call-and-response paradigm, the potential for intimacy and shared vocabulary. Good conversation requires sensitivity, unselfishness, and a balance between speaking and listening. This 'conversational' quality is an essential part of successful jamming: "If [conversation] doesn't happen, it's not good jazz" (ibid., p84).

The *conversation* of jamming takes place primarily through the musical 'utterances' of each musician, although linguistic and non-verbal modes of communication contribute as well. Musicians often refer to their musical contributions—the sound they make—as their *voice* (MacDonald, 2005, p404). In jamming, as in conversation, the musician receives direct feedback from their own instrument (in the form of their sound), and also near immediate feedback from the other musicians in the form of their musical response. In this tight loop, face-to-face co-location and instantaneous communication are all necessary for conversation (Urry, 2004). Each musician is a participant in some aspect of the conversation, and their influence and medium for this communication is their (musical) voice.

As with many metaphors, the analogue of jamming to spoken conversation can only be pushed so far. One key difference between jamming and spoken communication is the fact that the musicians are usually all 'speaking' at the same time,[2] whereas spoken conversation primarily proceeds in a turn-taking fashion. Additionally, music is far more ambiguous than language. While it is *possible* for a spoken sentence to be ambiguous, careful word choice and sentence structure can all but remove ambiguity. Because of this, it is usually relatively easy to assess whether a spoken contribution is an appropriate and helpful contribution to the conversation, or whether it is incorrect or irrelevant (Sawyer, 2004). A musical utterance, on the other hand, can be interpreted in many different ways, for example from a music-theoretic perspective (how well do the notes played fit within the current harmonic context) or from an emotional perspective (do the timbre and loudness of the sound fit the current mood). These factors mean that music is in a sense 'less constrained' than spoken communication—there are a greater number of coherent responses to a musical utterance than to a spoken one. The similarity lies in the fact that there is a need for coherence and relevance in the contributions of the participants.

Pinheiro also sees the musical interaction of jamming as a dialogue. Speaking from his experiences in the New York jazz scene:

> I define [a] jam session as a performance event in which any jazz musician can participate. It usually takes place weekly in jazz clubs after 9:30 pm, and it might last for several hours, until dawn. Using a 'core' repertoire as a starting point for improvisation, musicians develop musical 'dialogue' rooted in aesthetic principles that govern jazz performance. (Pinheiro, 2011)

In these situations the musicians will agree on a tune from the standard repertoire (these tunes are often referred to simply as 'standards') and take turns improvising. This improvisation usually proceeds according to a standard pattern. To begin, all musicians play the chorus (or *head*) of the tune, then the soloists take turns improvising while the rhythm section (drums, bass, and piano or guitar) keeps time and provides harmonic accompaniment. Finally, the group returns to the chorus one last time. There are many variations possible within this loose framework, and these variations are usually negotiated musically rather than verbally. There is certainly freedom to go 'off script', but in general there is a considerable amount of order to these jam sessions (Young and Matheson, 2000).

It may seem amazing to an outsider that musicians can assemble, possibly as complete strangers, and with the briefest of negotiations begin playing music with skill and coherence. This is possible because of the background of tunes and conventions shared between the participating musicians. These conventions are generally concerned with song-level (rather than note-level) aspects of the jam session—the order and number of verses and choruses rather than the specific notes played. In jazz, this 'note level' is where there is the most scope for improvisation and creativity.

[2]Jamming is an activity with *high synchronicity* (Dennis et al., 2008).

To continue the theme of jamming as conversation, the background knowledge is the lexicon and grammar of jamming—the musical 'atoms' which are the components of this language and the rules about how they can be combined. As Berliner (1994) points out, to achieve fluency in jamming requires acquiring a vocabulary, which takes many years to acquire. Each musician's vocabulary is a subset of the lexicon, which contains *all* of the words in a given language, or in this case all of the musical ideas and conventions of a jamming community.[3] To participate in the conversation of jamming in this community a musician must first learn the language.

The vocabulary of jamming also allows for allusion and quotation. Jazz drummer Ralph Peterson, when listening to a recording of himself, describes how one particular rhythmic exchange was a 'quotation' from a pair of famous standards.

> Yeah! 'Salt Peanuts' and 'Looney Tunes'—kind of a combination of the two. [Drummer] Art Blakey has a thing he plays. It's like: [he sings a rhythmic phrase from the song]. And [pianist] Geri played: [he sings Allen's standard response]. So I played the second half of the Art Blakey phrase: [he sings the second part of Blakey's drum pattern]. (Monson, 1997, p77)

The ubiquitous knowledge of the jazz canon amongst jazz musicians allows even small musical contributions to be understood and elaborated upon in the conversation of a jam session. These moments of co-ordination between musicians are a source of great pleasure for the musicians involved (Dempsey, 2008). A common background of tunes is essential for this to happen.

Another feature of the conversation of jamming is the *spontaneity* of the musical interaction. Despite the conventions and structure of jamming cultures, a jam session is not necessarily predictable or formulaic. While each member of the group brings their own experiences and sensibilities to the activity, the creative output of the group is not the singular vision of any of the individuals, or even the sum of their individual contributions:

> In collaborative improvisation, a creative product emerges that could not even in theory be created by an individual (Sawyer, 2007).

Sawyer sees in the jamming group the potential for what he calls *collaborative emergence* (Sawyer and DeZutter, 2009), for which he proposes four antecedents:

1. The activity has an unpredictable outcome, rather than a scripted, known endpoint;
2. There is moment-to-moment contingency: each person's action depends on the one just before;
3. The interactional effect of any given action can be changed by the subsequent actions of other participants; and

[3] In spoken language, different people have larger or smaller vocabularies, depending on the books they have read, their level of education, their memory for words, etc. There will be significant overlap in the vocabularies of two speakers of the same language; the more common will be in both vocabularies. The idea banks of different musicians will also have many things in common, such as musical scales or common chord voicings. Outside of this common ground, individual musicians will prefer certain phrasings or harmonic ideas over others, and this is one of the factors which gives rise to different musical utterances in a jam.

4. The process is collaborative, with each participant contributing equally.

This second antecedent—contingency—is particularly worth noting. Each musical utterance must be made in response to the current musical context, taking into account the contributions of all the other musicians and much of the conversation of jamming is dependent on what is happening 'in the moment'. Here again jamming bears a similarity to spoken conversation and what Garfinkel (1967) calls the 'awesome contingency of everyday life'. The jamming musician must both play and listen, act and react; balancing the desire to be fresh and original with the economies of falling back on familiar patterns and the need to fit musically with the other musicians.

This is not to say that jamming musicians do not think ahead, making long-range plans about the dynamics (soft-loud) of their playing or the gradual revelation of their musical ideas to the group. What it does mean is that there is no guarantee that these plans will come to fruition, because the other musicians are similarly making their own plans, and if these plans do not line up then the conversation of jamming will take the form of a negotiation, with each musician putting forward their musical vision but also responding to the ideas put forward by the other group members.

The reason Sawyer uses the term *emergence* here is because the resulting conversation of jamming is complex and dependent on so many factors. As noted by Pinheiro, even the acoustic quality of the performance space can have a determining influence on the jamming conversation:

> the acoustic quality of the performance space is also crucial for the improvisational process of the musicians in a jam session. The sound of his instrument varies from one club to another, even with the same equalisation and amplification system. This fact makes certain improvisational choices work in certain environments but not in others. (Pinheiro, 2011)

Here these subtle changes introduced by different environments are not irrelevant, they determine which choices *work*. Predicting ahead of time the success or effectiveness of a even a small musical contribution is very difficult, let alone the outcome of a whole jam session.

The third antecedent of jamming from the above list is perhaps the most interesting—that the contributions of each performer are ambiguous in the sense that their meaning (in context) depends not only on what precedes the contribution but on what *follows* it. The success of a particular musical contribution, such as a melody line or a rhythmic motif, can only be determined by the way it is integrated (or ignored) by the group. Individual contributions to the jamming group cannot be understood in isolation, order emerges only as the parts work together. This is a deeply interactionist picture of jamming, where meaning is made in the interaction between the musicians, and the success or failure of the jam depends on them *collaborating* successfully.

Jamming groups are also fluid; the actions and roles of the group members are not preordained but negotiated and re-negotiated on-the-fly. There is no pre-determined hierarchy of creativity and composition. This is in contrast to musical traditions where the composer composes a piece which is then performed by the musicians. In jamming (indeed in all improvisational music making) the musician and the composer are the same person, and 'composition time' and 'performance time' are collapsed into the singular moment of the jam.

There is no preordained order regarding how the roles of 'speaker' and 'listener' are filled, and indeed during the course of a jam the musicians may switch roles many times. This negotiation (and renegotiation) is part of the skill of jamming, so that all the necessary roles are filled at any one time.

This fluidity is even more pronounced in jamming traditions outside of jazz, so-called *free improvisation.* Tuedio (2006) describes the rhizomatic (non-hierarchical), self-organising nature of the music of the Grateful Dead. The Grateful Dead are a rock band renowned for their largely improvisational live shows, with extended instrumental interludes and significant variations between shows. Tuedio, like Sawyer, is committed to the idea that the best moments in a given jam are not pre-planned, and may indeed initially take the musicians by surprise. Because of the unpredictability of these magic moments, fans of the Grateful Dead would often follow the band around on tour, with the knowledge that each night's show will allow for different creative connections to be drawn, and that different musical structures will emerge each night.

This idea of 'saying something' is not meant to reduce jamming to a pure information exchange, there is real satisfaction to be found in having a *good* conversation:

> But that's what I like, when it feels like you've had a real good conversation. You know, and then people love it, people love that, I think. That's what jazz—you know, if you want to ask me what jazz is that's what it is, to me. You know, that's it. (sixth interviewee, quoted in MacDonald and Wilson, 2006, p65)

Jamming, like conversation, requires proficiency in the language, is contingent on moment to moment signals and decisions, and is decentralised. However, the description of jamming as conversation is not solely based on these nuts-and-bolts structural and interactional similarities. The similarity extends to the *feeling* of saying something. In the next section I shall further investigate the importance of this feeling as a driving force behind the jamming group.

3.2. The agitation of jamming

In discussing the second major facet of jamming in the literature—*agitation*—I am interested in what drives and catalyses the jamming group. The language of 'conversation' describes *how* jamming unfolds, while in agitation I am more interested in *why*. I have deliberately chosen to use the word agitation (whose Latin root *agitō* means *I move*) instead of motivation. This is because motivation is associated with need satisfaction, goal setting and other high-level cognitive functions. While these factors undoubtedly play a part in understanding why jamming happens, I am interested in the agitating forces at all levels: cognitive, emotional, and physiological. I am also interested not just in the *causes* but also in the catalysts—anything which has a driving influence, both in the moment-by-moment interaction in a given jam, but also in the lifetimes of a community of jamming musicians. For these reasons, the agitation of jamming is a more appropriate (and I believe more interesting) umbrella term. What animates the jamming group, and what makes it *move*?

The agitating forces which drive the activity of the jamming group fall into two broad categories: high-level factors which motivate and drive participation in jamming over several jam sessions (often over a lifetime), and low-level factors which are drive the jamming group forward during the act itself.

I shall address the ‘low-level agitation’ factors first. As I have discussed in section 3.1, Sawyer (and others) insist that the activity of jamming must be viewed at a group level—that each musician’s actions cannot be understood if examined in isolation. This is because a key driver of the activity of the jamming group is the interaction between musicians. Rather than being a series of spontaneous and unprompted outbursts of creativity, each musical utterance is a *response* to the current musical context. So one of the key agitating influences on the jamming group is the fact that *it is a group*—that there are multiple musicians who must be responsive to one another. Once the group has started jamming, this feedback loop of action and reaction can sustain it. Jamming groups are agitated endogenously, they are propelled by their own internal interaction rather than being driven by some outside force.

But what about each individual musician’s specific musical decisions about what sounds to make in the heat of the jam? As I have discussed in section 3.1, the broad structure of a jam session is often determined by convention. The opportunity for creativity in these jams lies in the particular notes, dynamics and articulations chosen by the musician in playing within these structures. Norgaard (2011) conducted an in-depth study of artist-level jazz musicians with a view to examining in detail the processes underlying the note-by-note choices made in the course of improvisation. In this study the musicians were given an opportunity to improvise over a standard tune. This improvisation was recorded, and immediately following the improvisation the musician was interviewed while listening to the recording phrase by phrase and encouraged to comment on their process for deciding which notes to play.

The musicians talked about using an ‘idea bank’ (again, here is the concept of a musical vocabulary) of known musical phrases, chord shapes or harmonic ideas. These ideas can be used one-off or incorporated into repeating motifs to add a sense of continuity and to build and release tension. The musical ideas in the idea bank were often slightly more general than the level of specific notes, such as ‘finish a phrase on the second scale degree’, or ‘use a sequence of descending arpeggios’. These generalities allow the musician to apply the idea in a wider variety of musician situations, filling in the specific notes in a way which is appropriate to the current musical context.

One experience which was consistently described by the musicians was the disconnect between the low-level note choices (which they were not necessarily conscious of) and the higher-level evaluative process of which they were conscious. The musicians described being ‘surprised’ on occasion by the sound coming out of their instrument, but rather than this being a problem this was often the catalyst for change or movement in their playing. As one of the musicians put it, “Hopefully something pops up that’s worth doing something with” (ibid.). In such cases the agitation—the source of new musical ideas—lies in the interaction between what the musician was planning and what is *actually played*, which is not always the same thing, due to sub-conscious processes and muscle memory. From the perspective of jamming as conversation, the utterances of individual musicians do not come forth fully formed—the musician is in dialogue even with themselves. This *internal* musical dialogue is also part of the agitation that guides the jam.

Unexpected sound events were not necessarily *mistakes* (although they may be). Norgaard describes it as the musicians’ ‘hands’ being ahead of their ‘brains’. The brain is working, it seems, at a higher level than the individual notes, making aesthetic decisions based on harmony, melody and dynamics, while appropriate notes and phrases are filtered from the idea bank in conformity with these aesthetic goals. Cochrane (2008) argues that even these decisions about the aesthetic and emotional dimensions of the music are not fully under the conscious control of the musician. Because music is an emotional medium rather than a purely informational one, the aesthetic decisions made by the musician are not determined by a dispassionate cognitive process, they are an expression of the emotional and affective state of the musician. Jamming is the result of the complex interplay between these conscious, pre-conscious and subconscious processes.

Bowers captures this idea well in his description of the experience of improvising in an electro-acoustic ensemble:

> I repeat a sharp blow watching my co-performers closely in case one of them is noticeably coordinating their production of the rubbing sound with my activity. Within the music, I am trying to investigate and diagnose the music. I am trying to find out what is making this sound by analysing a gesture which was involved in its production, all the while continuing to play and fold in my activity with that of the others. They give no sign of any gesture showing close synchrony with mine but, on another blow, one of the hand-held electric fans I have been using falls from the table to the floor. I notice that the rubber wheels at the base of the table are gently moving and that subject to my assault the whole table has lurched several centimetres towards the audience. It is these wheels against the Fylkingen cushioned floor which are making the friction sound. My contact mikes are picking up the vibration through the frame of the table. An unintended instrument has appeared. I move the table around, playing the rubbing sounds, varying their pitch, squeakiness and duration. Sten-Olof quickly finds a synthesiser patch derived from a physical model of friction sounds and the electro and the acoustic engage for a while. I bring the table back to its original position, quickly check my wiring, and look for something else to do. (Bowers, 2002, p1)

The unexpected sounds, the slip-ups and the miscommunications are an important agitating force in jamming. They catalyse strong reactions, both positive (euphoria) and negative (frustration). These moments represent the inflection points, the 'phase changes'; moments where the jamming group is knocked off its present course to explore new areas of music making.

Another driving force for the synchronised interaction and collaboration is entrainment between the jamming musicians. Entrainment is the "ability to coordinate the timing of our behaviours and rhythmically synchronise our attentional resources" (Gill, 2007), the unconscious synchronisation of behaviours, particularly rhythmic ones, which happen in group music making. Entrainment can manifest itself as tapping a foot along to a catchy rhythm or the rhythmic swaying and dancing of a crowd at a rock concert (Clayton et al., 2004).

In group music making, entrainment may manifest itself as synchronisation between head movements between musicians, as noted in a study of violin players by Varni (2008). There is no musical necessity that these head movements be synchronised. While it is difficult to play the violin with a still head, the head movement is secondary to the primary instrumental manipulation of the fingers, hands and arms. The synchrony—the shared movement—between the musicians is not deliberate, it is a consequence of the power of music and musical interaction over the whole musician.

This synchronised and rhythmic movement amongst musicians is sometimes called grooving (Doffman, 2009). The power of sound to catalyse and synchronise involuntary movement implies that the sounds that the musicians make are not inert, they have the capacity to move the listener (Cummins, 2009). Combined with the fact that in jamming (particularly in peak moments) the musicians in effect surrender conscious control of even their *own* sound, the complex behaviours which may emerge in the jamming group are clearly not within the conscious control of any musician. This power of sound to affect the listener is a feature which sets jamming apart from other group activities. A group of accountants collaborating on a spreadsheet is also a group activity, with improvisational dynamics as they discuss ideas about how best to balance the books. The difference between this scenario and jamming is that the spreadsheet is unlikely to engender the same emotional and physiological response in the accountants as the sound is capable of in the musicians. This 'high intensity' musical environment gives rise to the volatility and dynamism of jamming group.

Turning now to consider the higher-level agitating factors, I shall examine the forces which compel musicians to get (and stay) involved in jamming over long periods, indeed for their whole lives. What is it that keeps them coming back to jamming? While human behaviour is complicated and motivations are mixed, there is some literature addressing this question.

Flow (Csikszentmihalyi, 1991) is often used to describe 'peak experience' in jamming, and indeed music-making in general (Dubnov, 2005; MacDonald, Byrne, et al., 2006). Although Csikszentmihalyi was originally concerned with flow experiences in individuals, Sawyer (2006) described flow in improvisational groups such as jazz ensembles. Flow describes a scenario in which an individual's skill level is commensurate to the difficulty of the task being performed, most clearly demonstrated in the classic 'flow diagram' (see fig. 3.2). The intrinsic pleasure of finding flow in an activity provides an explanation for why some activities are inherently pleasurable and satisfying, even when they provide no discernible reward (outside of this satisfaction). Flow is a theory of *intrinsic motivation*, as distinct from the extrinsic rewards which often motivate participation in a given activity. There are ten elements which may contribute to a flow experience (Csikszentmihalyi and Rathunde, 1993, p60), although it is not necessary for all ten to be present in any given experience of flow.

1. Clear goals (expectations and rules are discernible and goals are attainable and align appropriately with one's skill set and abilities). Moreover, the challenge level and skill level should both be high.

2. Concentrating, a high degree of concentration on a limited field of attention (a person engaged in the activity will have the opportunity to focus and to delve deeply into it).

3. A loss of the feeling of self-consciousness, the merging of action and awareness. Action with awareness fades into action alone.

4. Distorted sense of time, one's subjective experience of time is altered.

5. Direct and immediate feedback (successes and failures in the course of the activity are apparent, so that behaviour can be adjusted as needed).

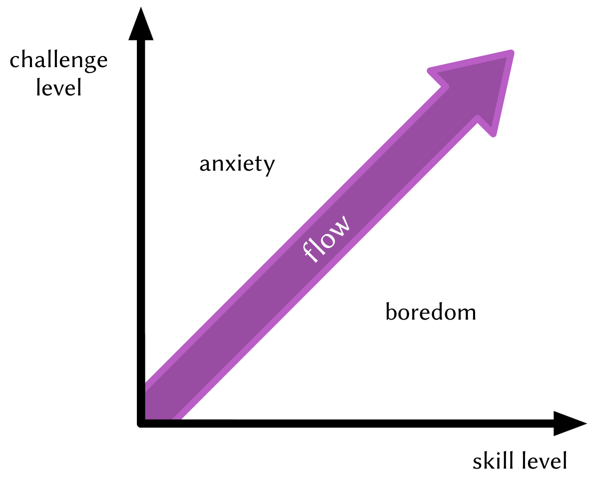


Figure 3.2.: Csikszentmihalyi's flow diagram. Finding flow requires a balance between the participant's skill and the challenge of the activity.

6. Balance between ability level and challenge (the activity is neither too easy nor too difficult).

7. A sense of personal control over the situation or activity.

8. The activity is intrinsically rewarding, so there is an effortlessness of action.

9. A lack of awareness of bodily needs (to the extent that one can reach a point of great hunger or fatigue without realising it)

10. Absorption into the activity, narrowing of the focus of awareness down to the activity itself, action awareness merging. Action with awareness fades into action alone.

Some of these features are clearly recognisable as features of jamming, such as the need for direct and immediate feedback. Furthermore, the general idea of flow occurring in the moments when conscious decision making gives way to complete absorption and 'surrender' to the task is one which is consistent with descriptions of jamming at its best:

> You know you always do your best soloing in the, fourth set of a gig when you're absolutely knackered and you don't care anymore and that's sorta when you, you kind of let go and then you kind of get that sort of...how'd you call it? Sort of a stream of consciousness, that's, you know your technique and all your practice and everything sort of, goes into autopilot and it lets it happen. (third interviewee, quoted in MacDonald and Wilson, 2006, p64)

These moments where the 'best soloing' happened when the musician was "absolutely knackered" (ibid.), when higher-level filtering processes had ceased to operate.

Other elements of flow do not map so neatly onto the types of peak experiences we see in jamming. For one thing, the first element of the list describes an activity with clear goals, but the open-ended and exploratory nature of a jam seems to preclude this condition being met. Similarly, the seventh element of flow is 'a sense of personal control'. Again, the conversational and emergent nature of jamming means that there is only limited personal control, each musician's contribution can only make sense in the context of the actions of all the other musicians. It is important to remember that the ten elements of flow listed above are not all necessary for a given flow experience.

The language of 'flow' does not come from the practitioners of jamming themselves. 'Grooving' or 'swinging' are more often used by musicians to describe the feeling of playing well together (Doffman, 2009). These terms have subtly different meanings depending on usage. They can refer to a specific beat or rhythmic pattern, or the practice of playing early on certain beats and late on others. They can also be used by musicians to refer to peak moments in a performance.

Jamming groups do not always reach these lofty peaks. One day a group might really be 'in the groove', the next day it might be flat. When it works, though, the experience of jamming together can provide a sense of satisfaction and connection with others that few other activities can (Mazzola, 2008). The sensation of being 'in the groove', while difficult to describe in words, represents a real shared experience prized by musicians across many different musical traditions (Lamont, 2009). These moments are not necessarily the moments where the music is being performed optimally, i.e. without mistakes. Instead, they refer to a subjective experience from the perspective of the musician:

> Optimal music performance, for most musicians, seems to involve reaching a state of consciousness outside of, and different from, that of the performer's ordinary experience. Although this state of consciousness often corresponds with the highest levels of execution, it is important to note that optimal music performance is not necessarily dependent on a particular type of music or the quality of its execution; it is not the optimal performance of music. Rather, the term refers to an unusually intense, heightened awareness, which for ease of identification, I refer to as the 'limit-experience.' (Knight, 2004, p1)

This 'limit' or 'peak' experience is not just a by-product produced as the jamming group seeks some other end, it is the ultimate reason for *why* musicians play. Alterhaug, drawing on the language of flow theory, puts it this way:

> All professional jazz musicians know that playing jazz—improvising—is about striving hard to obtain the ideal state in the 'golden moments'—ecstatic heights in musical interaction that are the *main reason for why we play.* But, unfortunately such moments occur all too seldom. However, behind the motivation and intention to reach a level of 'peak-performance' there has to be an existential urge. Besides such qualities, it is crucial that there is a good balance between challenges and skills. If this balance is not optimal, the musicians will either feel bored or anxious, and thus weaken their potential for interaction. This will have a negative impact on the ensemble's performance. However, when this balance is optimal, the musician will feel good; in this 'aesthetics of presence' a state is reached that is often referred to as flow. Musicians describe this state as 'being played' when they only observe their fingers playing their instrument. In this instance, the performer's condition is a kind of constructive uncertainty and confusion, thus being in a transcendental state. (Alterhaug, 2004, p105, emphasis added)

This is not only true for professional jazz musicians, either. Bloch conducted a series of interview with amateur musicians in garage rock bands on the experience of flow in group music making.

> "This sound simply stood up off the ground." His experience of flow was one of playing new chords in just the right musical spaces. They stopped after that because the rehearsal time was up. They felt exalted and happy. There were lots of jokes. He thinks of it as waves of energy. The feeling was a common one. Everyone felt that it had been a fantastic session. His reflection on this sequence of events is that it is only after rehearsal, when a number has been given form, that it becomes possible to "…speak the music. It's pure joy all the way through." (Bloch, 2000)

In summary, in this section I have discussed both the low and high-level factors which provide the impetus and agitation for the activity of jamming. At a low level, the use of 'idea banks', the positive influence of novelty (and even mistakes) and the power of sound and rhythm to set up a groove are all factors which drive the jamming group forward. At a higher level the experience of 'peak jamming' is the ultimate goal of the jamming group; it is what keeps the musicians coming back to jam sessions.

It is worth noting what factors *not* part of the agitation/motivation of jamming. The primary agitator is not financial remuneration, the adulation of an audience, or the preservation of a recorded artefact for posterity. Csikszentmihalyi uses the adjective *autotelic*[4] to describe any activity in which the feeling one gets when participating in the activity is the primary motivation for doing so. Whether this feeling is called 'flow', or 'being in the groove', or 'swinging', or not given a label at all: jamming musicians are chasing a feeling. Over long timescales, this is the primary agitating force behind the cultural practice of jamming.

3.3. The differentiation of jamming

The final facet of jamming that I shall discuss is the *differentiation* of jamming. By differentiation I mean the things which separate different jamming musicians and different jam sessions. Again, I shall be drawing on literature relating to improvisational group music making practices, and primarily jazz.

From a purely information-theoretic perspective, communication *requires* differentiation. Computing machines require the differentiation between the binary 1s and 0s on which they operate. In genetics the process of differentiation and mutation between genomes give rise to the diversity of organic life (Colebrook, 2003, p57). Conversation requires the differentiation of voices—imagine trying to have a group conversation without being able to determine who was speaking. This is akin to the experience of a native English speaker arriving in China (or vice versa) and being assaulted by a cacophony of voices speaking in a foreign tongue. While they may all be speaking their language perfectly well, to the foreigner the effect is one of being awash in a sea of strange sounds, unable to pick individual voices out of the crowd, unable to discern the segmentation of the speech into words and sentences.

[4] *auto* = self, *telos* = goal

At a practical level, then, each musician's sound must be distinct from all the other musician's sounds. This differentiation may be based on the fact that each musician is playing a different instrument, making differentiation easy. Even if each musician is playing the *same* instrument, though, the feedback from the instrument being played will allow the musician to pick their own sound out of the mix. The synchronisation between the movement of their hands and hearing the sound allows the brain to infer causality between the manipulation and its sonic effect.

The differentiation of musical voices is more than a mechanical necessity, though, it is a long term aspiration of the jamming musician. As discussed in section 3.1, jamming requires absorbing a corpus of tunes and conventions in order to participate in a meaningful way. The idea of having conventions and norms, though, connotes conformity and orthodoxy. How does a jamming musician stand out from the crowd?

> Assimilating the conventions of the genre such as rhythmic feel, articulation, scales, chords, compositional form, we are shown how jazz musicians seek to emerge from the shadows of their mentors to eventually blossom as individual creative artists. Crossing the bridge from imitating one's idols to musical originality is considered the goal. According to Berliner (1994, p276) it is only the handful of artists who manage this transition that produce 'compelling visions with major ramifications' for the music. (Lewis, 2006)

The jamming musician strives to find, amongst the background hum of genre conventions and musical influences, a distinctive voice as an individual artist. These are the artists whose own unique style then becomes part of the canon. The ability of the jamming musician to differentiate themselves is only possible because they know the conventions intimately, and the escape from the gravitational pull of these conventions is what *makes* a major artist. Differences between musicians are therefore not just a nuts-and-bolts matter of differences in repertoire and experience, but are the aspiration of the jamming musician.

> Jazz improvisation valorises subjectivity, emotion, the aesthetic, but also the openness and uncertainty that go against the fundamental goals of prediction and control so highly valued by the traditional sciences. A defining quality of creative improvisation is precisely the generation of the unpredictable, the unusual, the unforeseen, within the preexisting structures of the song form, navigating the edge between innovation and tradition (Berliner, 1994). (Montuori, 2003)

Jamming musicians may seek to stand out and differentiate themselves through their physical posture, dance moves or even their clothing.[5] A musician's primary vehicle for expression and differentiation, though, is their *sound*.

> One of the most important of these parameters is the quality of a musician's sound. [Jazz pianist] Aaron Goldberg states that the sound of a musician must strike the listener somehow and be convincing. "You know, I mean, does he make a beautiful sound from his instrument or if not a beautiful sound, at least a personal convincing sound? You know, that sticks in my head." (Pinheiro, 2011)

[5] "No change in musical style will survive unless it is accompanied by a change in clothing style. Rock is something to dress up to." –*Frank Zappa*

A musician's influence in the jam session is mediated via their sound, it is the manifestation of their agency. This influence at a base level is determined by which notes they play, but there is a great deal of room for subtlety.

> Many of the non-notable aspects of jazz improvisation—including tone colour, phrasing, dynamics, rhythmic coordination and intensity—as well as the intermusical connections that listeners hear in a jazz performance are among the seemingly ineffable physical qualities that produce emotional reactions in listeners. (Monson, 1997, p211)

The sound is powerful; it has the capacity to affect the listener at a cognitive level, but also at a base emotional level.

A musician's ability to differentiate and express themself through jamming is also tied to their identity (MacDonald, 2005). For some, jamming is the best form of expression available to them:

> that's one reason that I always come back to jazz because it's probably my best way, in everything, in life, of expressing myself. I always think that I'm especially bad with explaining myself with words, but give me a horn and a piece to play, and it's a completely different story. (p54 Dobson, 2010, musician J1, quoted in)

For this musician, the ability to express themselves is the reason they "always come back to jazz", the driver or the compelling force behind their jamming. The satisfaction of being able to express themselves more clearly in through his instrument than through any other avenue in their life (even language) is the thing which brings them back to jamming.

MacDonald (2005, p412) makes the point that this is a double edged sword, because there are also occasions where musicians may feel that they are not being allowed to express themselves as they would like. This may happen, for example, if a musician feels pigeon-holed and is forced into playing a style of music they are uncomfortable with or do not like. Similarly, any lack of fluency or inability to make the right sound in a given musical circumstance may be extremely frustrating and discouraging. Because their musical voice as jamming musicians is so deeply connected to their identity, an inability to differentiate or express oneself is not easily shrugged off. Differentiating between voices is an important day-to-day skill for the conversation of jamming, but in a broader sense differentiation and finding a unique voice is a long term goal for the jamming musician.

3.4. The digitisation of jamming

Jamming, like almost all cultural practices, has not escaped the influence of computing devices. I have thus far spoken only of jamming in the context of traditional instruments made out of wood, skin and metal. Finally in this chapter I shall consider the impact that digital technologies can have (and are already having) on jamming.

Musicians and artists have historically been amongst the avant garde of technology adoption, for example in the early use of computers in music making (Mathews, 1963). The use of computers can take many forms, from music digital audio recording software on desktop computers to the creation of custom 'instruments' for musical expression which incorporate

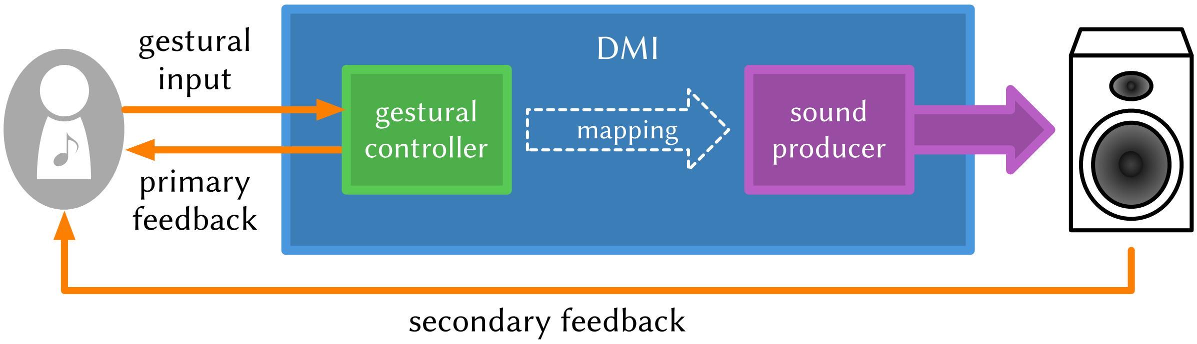


Figure 3.3.: A DMI has two main components: a *gestural controller*, which registers the physical manipulations of the musician, and a *sound producer*, which produces the sound (adapted from Miranda and Wanderley, 2006). 'In between' these components is a *mapping* from gestures into sound. The mapping from gestures into sound is a fertile area for innovation in DMI design. The musician receives two kinds of feedback—primary feedback (tactile, proprioceptive) from their physical interaction with the instrument, and secondary (auditory) feedback in the form of the sounds produced by the instrument.

digital technologies. Digital Musical Instrument (DMI)s can take many different forms (see Paine (2010) for a taxonomy of DMI design approaches). These DMIs afford musicians new opportunities for jamming, because they can be personal—a specific tangible object to own, use and master.

The general structure of a DMI is shown in fig. 3.3. DMIs need not be designed specifically for jamming. The instrument need not be played improvisationally (the music may be fully notated), or it may be designed to be played individually rather than in a group. However, there is certainly a clear opportunity for providing computational support to the cultural practice of jamming. The NIME community (which was spun out of the CHI conference in 2001) is motivated by the question

> how can we create new interfaces to play computers in a way that is appropriate to human brains and bodies? (Fels and Lyons, 2009)

This community is principally concerned with the using computers for musical *expression* (Dobrian and Koppelman, 2006), a goal very much in accordance with Monson's (1997) dictum of jamming as 'saying something'. In this research there is a constant tension between the artistic and the HCI agendas, particularly in regard to evaluation (O'Modhrain, 2011).

Blaine and Fels (2003b) have discussed the inherent tension between accessibility and expressivity in DMI design, particularly in collaborative musical contexts.

> Designers of collaborative devices that are easy to control but have limited expressive capabilities are challenged not only to conceive of opportunities for musical exploration, but must also cultivate meaningful social interactions and experiences for the players. In a collaborative musical environment, it becomes even more imperative that the technology serves primarily as a catalyst for social interaction, rather than as the focus of

> the experience (Robson, 2001). Conversely, interfaces that have extended expressive capabilities tend to be more difficult to control and cater more to the expert player. For designers of most musical interfaces, the overriding challenge is to strike a balance of multimodal interaction using discrete and continuous controls (Tanaka, 2002; Verplank and Sapp, 2001), and generally, limit rather than increase the number of features and opportunities for creativity (Cook, 2001). (Blaine and Fels, 2003b)

The point about limiting rather than increasing features is a particularly interesting one. In jamming, it would seems that expressivity is more important than accessibility. The ability of the musicians to converse musically in a jam is dependent on their ability to articulate their musical ideas in an immediate and responsive fashion. Still, expressivity is not the same thing as complexity, and making an interface more expressive is not as simple as adding more buttons and knobs to control more parameters. Constraint is not inherently prohibitive from a creative standpoint. Gurevich et al. (2010) developed a 'one button instrument': a battery-powered box with a small loudspeaker and a single button which triggered a (morse code style) beep. There was no pitch, timbre or volume control, only an on/off trigger. Still, in only a week of practising (solo jamming), the musicians developed innovative ways to play the instrument, such as filtering the sound by covering the speaker with a hand, and even drumming on the box itself. It is at least important to be aware in designing a DMI of this accessibility-expressivity tension, and I shall address this tension in the 'mapping' section of each Viscotheque chapter (sections 4.2, 5.2 and 6.2).

Evaluation techniques from more traditional HCI have been adapted for musical interaction contexts such as getting participants to complete basic musical tasks (Wanderley and Orio, 2002) which are comparatively easy to assess. As an open-ended activity, though, jamming is not amenable to this type of reductionism. 'Mistakes' such as wrong notes are often sites of inspiration, perturbing the musical status quo and having an overall positive effect on the trajectory of a musical performance (McDermott et al., 2011).

The potential of the mobile phone as a musical instrument has been noticed by the NIME community (Gaye et al., 2006). Modern 'smartphones', with their capacitive multi-touch screens and array of other sensors (Essl and Rohs, 2009), have proven popular for creating new instruments and contexts for group music making (Dahl and Wang, 2010; Halpern et al., 2011; Oh et al., 2010; Zhou et al., 2011).[6]

The affordability and availability of smartphones gives them an advantage over custom hardware. From a design perspective this may seem limiting—there is clearly more creative freedom afforded the designer when the DMI is built from scratch. However, the smartphone's familiarity and ubiquity can be exploited to enable increased access to the experience of jamming (Blaine and Fels, 2003a). In this technological shift, musicians are finding new ways to jam together, and to share in that familiar collaborative, improvisational jamming experience (Tanaka, 2006).

The shift to digital instruments opens up new possibilities for jamming in two ways. Firstly, the DMI is not limited in the sounds it can produce by its physical and acoustic properties.

[6]It should also be noted that there is a great deal of exciting DMI research going on outside the academy. Websites like *Create Digital Music* and *SoundCloud* are excellent showcases of the possibilities of computers in music making.

This freedom is not unlike the opportunities created by the development of electric amplification and electronic processing in 20th century popular music. This expanded the sonic potential of jazz musicians of the time, as can be seen in the use of delay and other effects by Jaco Pastorius and Miles Davis (Nesbitt, 2010) or the studio production techniques of Pink Floyd and Brian Eno.

DMIs *may* be made to reproduce the sounds of traditional instruments, for instance in the case of a digital keyboard which responds to key presses by playing back pre-recorded samples from an acoustic piano. More interestingly, though, DMIs may map input gestures into sounds *unlike* any produced by acoustic instruments. A DMI is not completely freed from the realities of physical sound reproduction, some sort of physical loudspeaker is required to transform the digital data stream produced by the sound generator into vibrations in the air which can be heard. Still, the move to digital sound sources and synthesis represents new sonic possibilities for the jamming (and indeed any) musician.

The second way that DMIs open up new potentials for jamming is due to their novelty. The smartphone, as an instrument, does not have the hundreds of years of musical tradition and repertoire that are associated with the violin. Designers of DMIs *can* choose to tie their designs into these traditions and the musical understandings and sensibilities of the musicians who play them. The opportunity exists, however, to break free from notions of tonality, expectations of which notes can follow each other, and other constraints on the way the jamming group can act musically.

Borgmann (1987) is careful to point out that technology does not necessarily make things 'better', and HCI practitioners must be careful when wading into the domain of ethics. An optimistic reading of the development of accessible, smartphone-based DMIs may consider it a 'democratisation' (Tanaka, 2010) of music-making. The experience of jamming is being brought within the reach of anyone with an appropriate phone in their pocket. The nature of a phone as a constant companion also opens up the possibility of spontaneous jam sessions, turning idle moments and new acquaintances into opportunities to jam. A more pessimistic interpretation of this trend may lament the dumbing-down of a complex, skillful activity, and perhaps a loss of the nuance and ceremony surrounding jamming. The truth probably lies somewhere between these two poles, but it is important to remember that this next chapter of ubiquitous digital musical interaction has not yet been played out.

3.5. Chapter summary

It may seem as though I am using the term 'jamming' injudiciously, or that I expand and contract the definition given earlier to include or exclude different examples from the literature as suits my purposes. I am not trying to draw a sharp line around that which is and which is not jamming; rather I am trying to convince the reader that such music making contexts *exist*, and shed some light on their essential characteristics.

In jamming, the three facets of conversation, agitation and differentiation are crucial to understanding the high and low level mechanics of the jamming group. The three aspects are not cleanly differentiated; they are interwoven and overlapping. Coarsely, it may be said that the low-level *conversational* interaction happens in pursuit of a mid-level *agitation* towards the feeling of peak experience, while the high-level lifetime goal of the jamming musician is to *differentiate* themselves and find a sound all their own. Although this glosses over many of the subtleties discussed in this chapter, it may be helpful as an indicator of the broad picture of how these facets relate to one another.

The rationale for providing this breakdown is to capture the key features of jamming—what it is, how it works and why it happens. There are three points of overlap between the practice of jamming and the commitments of third wave HCI research:

- Jamming is a complex activity which cannot be reduced to the sum of its parts.

- The highest motivation and the greatest reward is the *experience* of participating in a good jam.

- The goal of the jamming musician is to differentiate oneself (to find a unique voice) and such differences should be celebrated.

The importance of conversation, the motivating power of chasing the *feeling* of a great jam, and the desire to find a unique voice and make a real contribution are still relevant as jamming enters a digital age. This is helpful to keep in mind as we seek to design interactive digital artefacts to support this experience. My purposes for this will become clear in the next chapter as I discuss *Viscotheque*, a computer-supported environment for jamming.

4. Viscotheque v1

Version 1 of the Viscotheque was developed in early 2009 and the inaugural performance of the system was on April 8, 2009. The Viscotheque v1 jam session took the form of a virtual drum circle. Each distributed performer was assigned control of a certain type of drum, either a djembe, taiko or conga. The inspiration for this design was from drum circles where members of a circle participate in poly-rhythmic improvisational music making, which are popular amongst hippie and neo-pagan subcultures (Telesco and Waterhawk, 2003).

Some of the work presented in this chapter was presented in the long paper B. Swift, H. J. Gardner, and A. Riddell (2009). "Distributed Performance in Live Coding". In: *ACMC '09: Proceedings of the Australasian Computer Music Conference.*

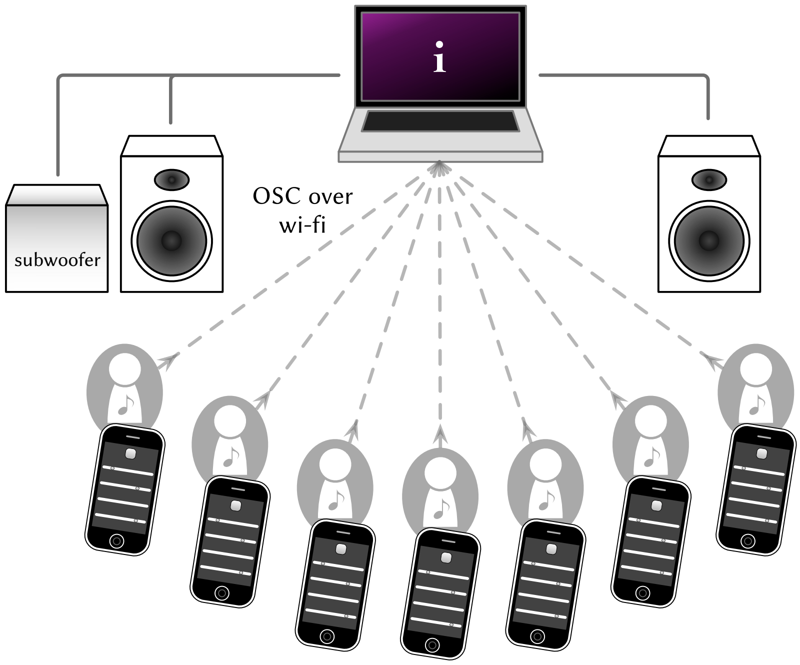
4.1. Architecture

Viscotheque v1 used a client-server architecture, with a thin-client iOS application running on the iPhone and a Viscotheque server running on a laptop on the Local Area Network (LAN). Each iPhone sends OSC (Wright, 2005) messages to the server over the WLAN. These control messages were interpreted by the server to trigger the appropriate sound, and all the musician's sounds are played back through a pair of loudspeakers. This process is shown in fig. 4.1(b).

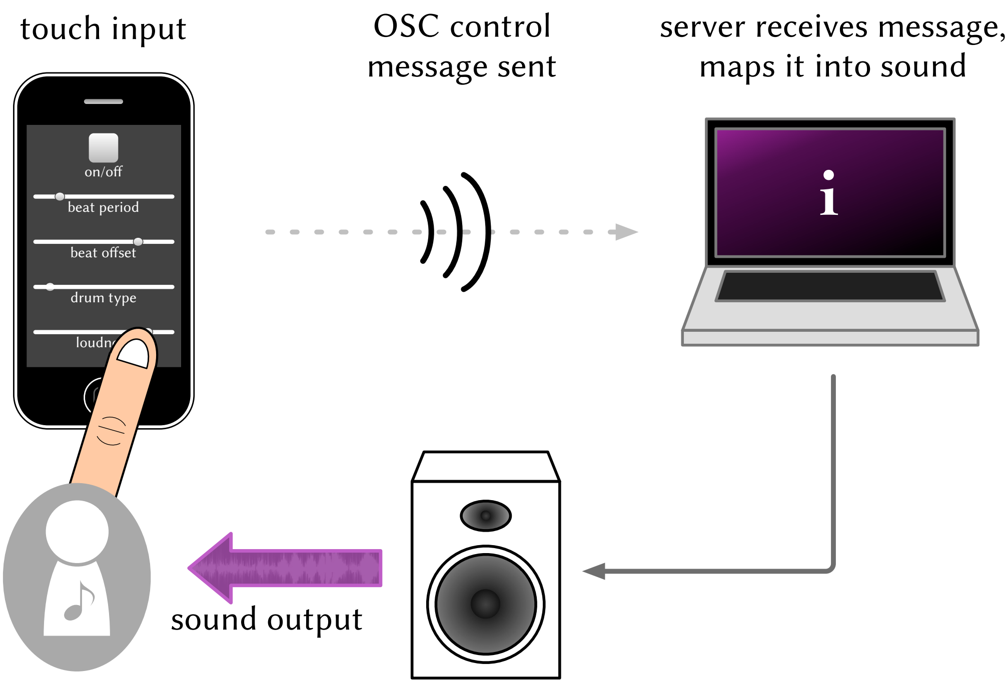
The Viscotheque was designed for co-located musicians, all jamming together in the same room. While there is no technical reason to impose this restriction, peak moments in jamming are shared experiences, and non-verbal and bodily communication is an important part of this activity. A drum circle affords the participant the experience of being part of a sound-

Hardware	Qty.	Details
MacBook Pro (server)	1	2.4GHz, 2MB RAM, Mac OS X 10.6
iOS device (client)	1+	this is the 'instrument' that the musicians play 1 iOS device per musician
Loudspeakers	2	KRK Rokit 6 full range monitor speakers, 68W RMS, frequency response 48Hz–20kHz (front of house)
	1	KRK RP10S subwoofer 150W RMS
Wi-fi router	1	Hosting the WLAN network

Table 4.1.: The hardware requirements for Viscotheque v1. The hardware for the system were largely the same between revisions, the main changes in version 2 and 3 were in the input$\rightarrow$output mapping and the experimental protocol.



(a) Viscotheque v1 system architecture. The musicians are co-located and seated in a semi-circular configuration. All sonic output comes through a pair of stereo speakers in the jam room. The Viscotheque server running on the laptop is implemented in the impromptu audiovisual programming environment.



(b) Viscotheque data flow. A finger touch on the interface is transmitted as an OSC message, received and interpreted by the server, and the appropriate sound generated (or existing sound modified) based on the gesture→sound output mapping.

Figure 4.1.: The system architecture and input→output interaction loop for Viscotheque v1.

producing collective, and the Viscotheque was designed to retain this sense of being part of a larger music making group. The system was capable of flexibly adding and removing devices on the fly—there was no hard limit to the number of musicians who could participate. The complete hardware requirements for the system are shown in table 4.1.

4.1.1. Client-server architecture

These two software applications (the client and the server) implement the mapping between input touches and output sound for each musician. This mapping is the heart of the Viscotheque instrument, and is described in detail in the next section (section 4.2). Although the mobile devices cannot be used without the server, we shall often refer to the iPhone running the Viscotheque application as the 'instrument'. From the musician's perspective the device *is* the instrument, it is the artefact by which their physical manipulations are transformed into sound. I shall often use the term 'device' to refer to the iOS device + client application, but it is important to remember that the Viscotheque instrument is the combination of software, hardware and sound reproduction infrastructure.

The iPhone (and indeed any equivalent 'smartphone', such as an Android device) possesses all the required hardware to be a standalone instrument—it has a range of sensors which could be used for input, and it has a speaker and audio jack for output. However, there were two reasons why a client-server architecture was chosen for Viscotheque:

Shared audio output
The point of jamming is to play *together*, and this requires that each musician can hear all the other musicians as well as themselves. Having each musician listen to their sound only (say, through their a pair of headphones) is not sufficient. While the iPhone does have a loudspeaker, it is small and incapable of the volume and bass response necessary for a satisfying jam. With a central server, the sound can be played for all musicians through a set of full range speakers connected to the server, which are capable of satisfactory loudness and fidelity.

It would have been technically possible to take each musician's audio stream and send it digitally to each other musician to be mixed into their own audio stream (perhaps at a softer level than their own contributions). However, without a central server each device would be required to transmit its sound to *every* other device in a peer-to-peer fashion, requiring n^2 connections for n devices. This would have limited the system's ability to scale to large numbers of devices., as well as being more complicated and susceptible to network latency issues. With a central server, only n connections are required—each device connects only to the server.

The key disadvantage of the 'shared sound output only' setup is the lack of an individual monitor output for each musician where they can listen to their sound in isolation. This is an issue with any electric instrument as well—any instrument in which the sound amplification production is decoupled from the physical instrument. This problem of identifying 'who makes what sound' (Merritt et al., 2010) is a tricky one, and many of the refinements to the Viscotheque system in subsequent versions (particularly in adding visual feedback) were made in response to this issue.

Logging
The other key benefit of having a central server is the ease with which the whole jam session can be logged for later analysis. Because there is one canonical (stereo) output audio stream, it is simple to record this audio data, and it is a faithful representation of the sound produced and heard by the musicians during the jam. If each device produces its own sound output, then these output streams must be mixed and synchronised in the final recording, and this is difficult given the sensitivity of audio data to temporal jitter.

It is a limitation of the Viscotheque server that only the *aggregate* stereo output mix can be recorded—it is not possible to capture a direct feed for each musician individually. This is because of the way that the audio recording infrastructure works in Impromptu.

It is not just the audio data which can be logged—all the raw touch input and interaction data can be recorded as well. This is useful for visualising each musician's patterns of interaction with the device, as shown in section 4.4. If the touch interaction data was logged individually on each device then it would be necessary to synchronise the time-stamps between the devices so that each musician's log data could be interleaved to produce a log of the whole jam session. Again, this is not impossible, but this was also an influence in deciding to use a client-server architecture for Viscotheque.

4.1.2. Viscotheque server

The Viscotheque server was a program written for the *Impromptu* media arts programming environment (Sorensen and Gardner, 2010) which runs on Mac OS X. Impromptu is essentially an interpreter and runtime for the *Scheme* programming language (Sussman and Steele, 1975) with strict timing semantics. Function calls in Impromptu may be scheduled with sample-accurate (from the perspective of the 44.1kHz audio output sample rate) precision, which means that rhythmic material can be scheduled easily, and that musical or visual events which are supposed to happen simultaneously are guaranteed to do so. On top of this, Impromptu is an *AudioUnit* host, allowing any audio plugin which conforms to the AudioUnit specification (*Audio Unit Programming Guide* 2007) to be used for sound generation. Impromptu also contains bindings for the Quartz 2D and OpenGL graphics libraries. I chose to build the Viscotheque server in Impromptu because of this built-in support for sound generation, OSC handling and audio recording and logging.

All of audio data manipulation in the system happened server-side, which allowed the Viscotheque iPhone app to be a 'thin' client, only transmitting control data (rather than audio data) over the network (the control message format is shown in fig. 4.2). This allowed the OSC messages to be sent using the User Datagram Protocol (UDP) protocol, which is a low-latency connectionless data transfer protocol (Postel, 1980). This was done to allow the interface to be as responsive as possible, so that any gestural input is reflected in the output sound as close to instantaneously as possible. Unlike Transmission Control Protocol (TCP), which re-transmits dropped packets and guarantees delivery at the cost of higher potential latency and reduced throughput, UDP is a simpler 'fire-and-forget' protocol. Packets are not guaranteed to arrive in order, they are simply routed to the destination Internet Protocol (IP) address as quickly as possible. This would be a problem if the audio data itself were being

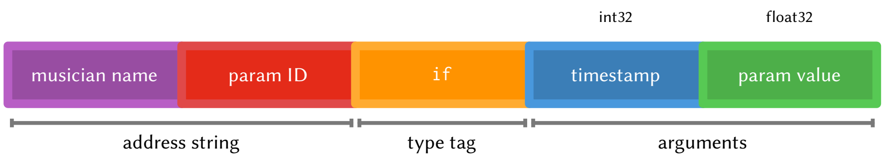


Figure 4.2.: Viscotheque v1 OSC control message structure. These messages were generated by the client and sent to the server to be mapped into sound (and also dumped to a log file for later analysis). The param ID portion of the address string indicated which slider widget (and therefore which parameter) the value belonged to (see section 4.2).

transmitted, because lost or jumbled packets would result in pops and clicks in the audio track. When only the control messages are sent, then the sound output is at least error-free, even if the musician's intentions may be not reflected perfectly in the output sound. Minimal latency rather than increased bandwidth was the primary reason for using UDP in Viscotheque.

The speaker setup for Viscotheque v1 was a stereo pair of KRK Rokit 6[1] full-range monitor speakers for 'front of house' plus a KRK RP10S[2] subwoofer (see table 4.1 for details). This setup is capable of producing an accurate sound field at moderate to loud volumes in a small-to-medium sized room. The speakers were connected to a 2008 MacBook Pro laptop running Impromptu and the Viscotheque server application. The hardware audio interface was a MOTU Ultralite Mk3 FireWire. The output signal was stereo only (there was no discrete low frequency channel for the subwoofer) so the front of house speakers were connected through the subwoofer. The KRK RP10S subwoofer contains a crossover which allows it to play the low-frequency content while passing the high-frequency content through unchanged to the Rokit 6s.

4.1.3. Viscotheque client

In April 2009, at the time of the Viscotheque v1 jam session, the iOS hardware platform included the iPhone (the current model the iPhone 3G) and the iPod Touch. The iPad had not yet been released. The iOS platform was chosen above other competing smartphone platforms (e.g. Google's Android or Nokia's Symbian) for three main reasons:

Popularity
Amongst local, known musicians there were a higher number of iPhone users than users of any other platform. Admittedly this was anecdotal evidence, but the plan was to first test the instrument in the local music community.

Familiarity
As a developer, I was already familiar with the iOS developer tools and application frame-

[1] `http://www.krksys.com/krk-studio-monitor-speakers/rokit/rokit-6.html`
[2] `http://www.krksys.com/krk-subwoofers/10s.html`

works. While this was less of an issue for version 1 (with the third-party MRMR app), it made the development of a custom Viscotheque application in versions 2 and 3 much easier.

Accessibility
The iOS App Store provided a easy way for musicians to get the iOS app. While this 'one canonical distribution channel' approach proved problematic when combined with the use of a third-party iOS application (see section 5.1.2), it was a key part of the 'pick up and play' accessibility of the Viscotheque as an instrument for jamming.

In version 1 of the Viscotheque instrument, the touchscreen was the sole gestural control input. The touchscreen was exactly the same between the iPhone and the iPod Touch—a 3.5" diagonal capacitive touch screen[3] capable of tracking up to 5 simultaneous touches.

For version 1 of the Viscotheque system, the open-source MRMR (*MRMR - Open Mobile Touch Protocol*) iOS application was used to provide the interface and OSC client running on the device. This application was a (free) download from the iPhone App Store, and allowed for a customisable interface of buttons, 1D sliders and 2D 'touch zones' to be presented to the musician. The interfaces were specified using a custom grid-based layout scheme called the MRMR protocol.[4] The interface used for the Viscotheque v1 drum circle is shown in fig. 4.3(b). These interfaces are specified and set up on the running MRMR app by sending the device a string representation of the interface (as per the MRMR interface protocol) over a TCP connection. Any system capable of making a TCP connection and receiving OSC messages is therefore capable of acting as a server for the MRMR client app.

Using the MRMR app made it easy for a musician with an iOS device to participate in the jam session. From start to finish, there were five simple steps to participating in the jam session:

1. Download the MRMR app from the App Store.
2. Join the Viscotheque WLAN. This is the network the server is connected to.
3. Launch the Viscotheque app on the iOS device.
4. In the settings tab (see fig. 4.3(a)), enter a name and the IP address of the Viscotheque server (which was clearly displayed on a screen in the jamming room).
5. Start 'playing' the instrument using the touch interface on the device's screen. From this point, the musician's sound could be heard through the shared loudspeakers.

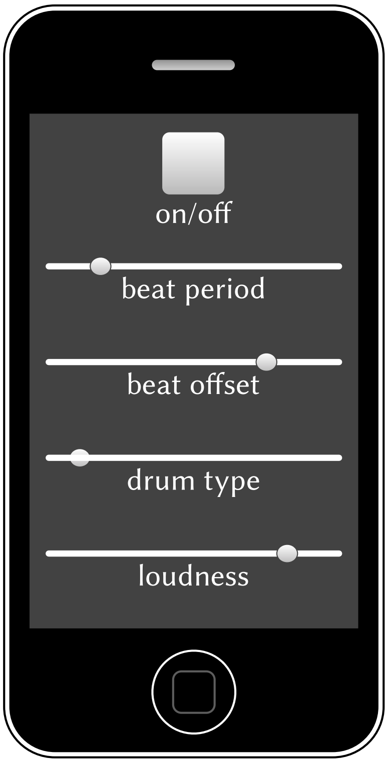
4.2. Mapping

The sound mapping for the Viscotheque v1 DMI was the key aspect of the instrument (see fig. 3.3). It was through this mapping that the musicians control their sound—their vehicle for conversation, agitation and differentiation.

[3] Specifications are available at `http://http://support.apple.com/kb/SP495`

[4] The MRMR interface protocol is documented at `http://mrmr.noisepages.com/mrmr-interface-protocol/`

(a) The MRMR app welcome screen. Upon start-up, the musician would need to navigate to the 'Prefs' (preferences) tab and enter their name and the IP address of the Viscotheque server, which was displayed on a screen in the jamming room. After this was done, the server would send the interface in fig. 4.3(b) to the musician and they were able to start playing.



(b) Viscotheque v1 app interface. The parameters exposed to each musician were for controlling a single periodic drum tap. The idea was that each musician could only control a single 'pulse train' of drum hits. As they jammed together, the sum of all these pulse trains would therefore be a complex poly-rhythm of with different accents and timbres. The effect that each parameter had on the drum output is described in table 4.2.

Figure 4.3.: Viscotheque app version 1 home screen and music making interface.

Fundamentally, the Viscotheque v1 DMI was a drum—or at least it produced drum sounds. However, the mapping between gestural input and sound output was not as direct as a traditional drum, where each strike of the skin produces an immediate sound. Rather, each musician controlled a regular 'pulse train' of drum hits—a metronomic sequence of drum sounds with constant tempo. Each musician's drum would continue to beat even without any input on their behalf. The Viscotheque instrument was like a fancy metronome, where instead of only controlling the tempo the musicians were able to control four aspects of the sound (as well as being able to turn the loop on and off).

The control parameters are shown in table 4.2. Each of these parameters were manipulable by the musician at any time, using a toggle button widget[5] for on/off or a one dimensional slider widget for the other four parameters. In manipulating these widgets the musicians played the instrument, either varying the parameters quickly to produce less rhythmic or consistent patterns, or leaving the parameters mostly constant to produce consistent, stable drum patterns. This process may be difficult to visualise from a textual description, so a visual representation is shown in fig. 4.4.

The term 'mapping' implies a domain and a range: a mapping *from* somewhere *to* somewhere. A mapping defines a relationship between an input space and an output space. In DMI design these are not necessarily *vector spaces* in the mathematical sense, it is a more general relationship between the what the musician *does* (their input) and the sound that they make (their output). The input and output spaces for the Viscotheque v1 mapping were:

Input space
The input space is the space of the parameters the musician can manipulate to change the sound produced by the instrument. In the case of Viscotheque v1, this space had five dimensions: the four sliders plus the on/off button (see table 4.2). Mathematically, this is the set $x \in [0,1]^4 \times \{0,1\}$ which is not strictly a vector space because it is not closed under addition. The musician through their playing controlled the trajectory through this input space.

Output space
The output space is the collection of all the possible sounds the instrument can produce. In practice, this was the set of all possible drum patterns the instrument could produce. While it is difficult to determine the dimension of this space in a meaningful way,[6] it is certainly orders of magnitude larger than the input space.

Each musician controlled their own sonic output, one musician could not affect the sound of any other. Each musician's sound was mixed together and played through the same set of speakers, the musicians did not have an individual 'fold-back' speaker to monitor their own contributions in isolation. The musicians therefore had to take care to listen to one another, and not to simply make the loudest noise possible and drown one another out.

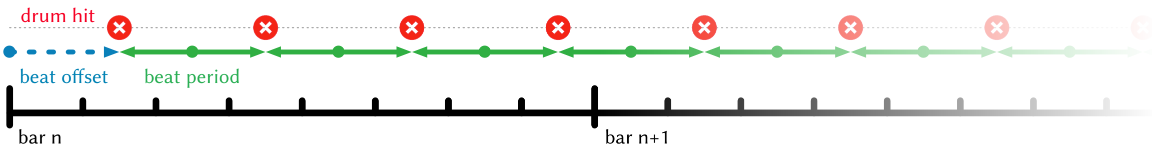
The Viscotheque interface was designed for real-time interaction. As discussed in chapter 3, the interplay of improvisational music-making requires instantaneous choices to be made about how to contribute musically to the overall sound at any given time. For this

[5]the term *widget* refers to any interface element, either a button or a slider

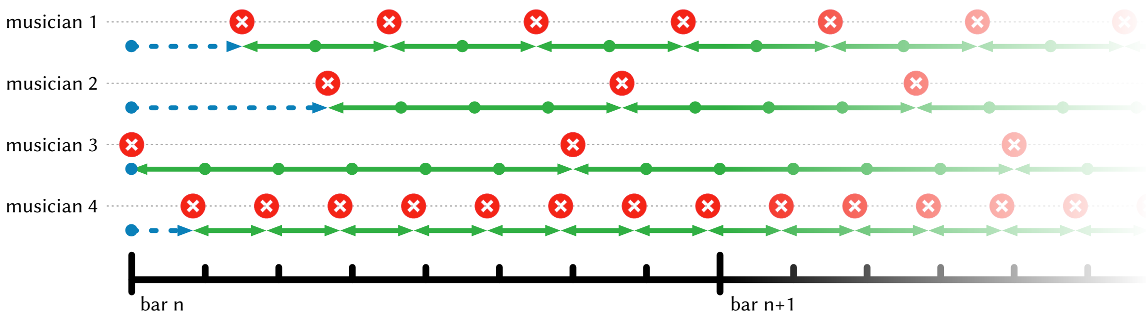
[6]with digital audio at a bit depth of 16 bit and at a sample rate of 44100Hz, the theoretical number of different output streams of length S seconds is $2^{16\times 44100\times S}$

Parameter	Description
On/off	Toggle the drum ‘on’ or ‘off’.
Beat period	The ‘tempo’ of the drum metronome. The values taken by this parameter were quantised to 32 equally spaced points between the smallest between-hit duration (0.5 seconds, which represents one ‘sixteenth note’ at 120bpm) and the largest between-hit duration (4 seconds, which represents 2 whole bars at 120bpm).
Beat offset	The ‘phase’ term—the relative offset of the metronome from the global beginning of the bar. Because there was no indicator of when the beginning of the bar was (unless one of the musicians was playing their drum with offset = 0), the offset parameter was largely used to adjust by each musician to adjust their drum hits relative to one another. This value was not quantised, and could take any value up to the current value of the ‘beat period’ parameter.
Drum type	Select from five available drum sounds: timpani (a deep, bassy drum), conga ‘heel’ strike, conga ‘slap’ (a high pitched, cutting drum hit), taiko (traditional Japanese) drum and a djembe. All of the drum sounds were professionally recorded samples taken from the Native Instruments *Battery 3* sound library.
Loudness	Loudness of the drum playback.

Table 4.2.: The control parameters for the Viscotheque v1 drum circle instrument.



(a) An example drum pattern (for one musician). The beat offset into the bar can take any value less than the beat period, while the beat period (the time between drum hits) is quantised with 32 steps between the shortest and longest possible periods (although only eight quantisation steps are shown in the figure for clarity). The pattern will loop like this until the parameters are changed by the musician.



(b) Each musician controls just one drum, but the total output of the jamming group is the sum of all these contributions. The drum patterns show here are static; this is a period where the musicians do not alter the parameters. However, the musicians are free to modify all of these parameters in real-time, and these alterations are reflected in their sound immediately.

Figure 4.4.: The 'drum circle' control parameters. In addition to the beat offset and beat period controls, each musician can also control the type of drum they are playing (i.e. the timbre of their sound) and their volume. Finally, the musicians can turn their drum on and off. Although each musician's parameter space is limited, complex poly-rhythms may emerge from their interaction.

reason, any finger touch on the screen had an immediate (although potentially subtle) effect. This closed the feedback loop between the musician and the environment, allowing them to explore the extent of their sonic agency.

Flores et al. (2010) proposed some mobile music interaction 'design patterns'—mapping strategies and categorisations specific to music making with mobile devices. They proposed four patterns: natural interaction, event sequencing, process control, and sound mixing. In regard to these, Viscotheque v1 is a *process control* interface. This design decision was influenced by a desire to reign in the potential chaos of multiple musicians, as providing limited control to the musicians can control can offer "more cohesive sound spaces in multiplayer environments" (Blaine and Fels, 2003b).

The musician's gestures were not directly mapped into sound in a direct 'touch→sound' fashion.[7] Rather, the musicians controlled a generative musical process (in this case a drum pattern) via a limited number of parameters. The musician was jamming at a higher level of abstraction, their creative decisions were constrained to their trajectory through the low-dimensional 'control parameter space' exposed to them by the instrument.

The Viscotheque was designed to be simple; it was simple to join a jam, simple to make a sound, and therefore simple to begin *interacting* with the other jamming musicians—making sounds which fit with, and contribute to, the current musical context. The jamming musicians would be learning the mapping on the fly, learning the sonic effects of manipulating the different parameters at their disposal. Using only five control parameters with a *process control* mapping meant that the interface was *accessible* at the expense of being less *expressive*.

This 'accessibility above expressivity' approach may seem at odds with the emphasis in chapter 3 on the importance of assimilating the conventions and norms of jamming. However, as discussed in section 3.4, constraint is a natural part of any instrumental design, and even extreme constraints need not stifle creativity in music making.

The other reason that a low-dimensional mapping was used was to push the musicians to interact and play *together*. In a real-life drum circle, the rhythmic complexity of the drumming does not come from any one musician in isolation but from the layering of many different rhythms, drums and accenting patterns. The Viscotheque mapping similarly enforced this 'one musician, one sound' idea, with a view to encouraging creative interaction to create more complex rhythmic patterns.

To recap, in Viscotheque v1 each musician controlled a metronome-like generative drum process using one-dimensional sliders to adjust four simple parameters. The interface was deliberately simple so that the musicians could easily join in and start interacting with each other. The total effect was that of a multi-musician drum circle, with each musician contributing a simple rhythm which, when overlaid upon all the others, would hopefully lead to poly-rhythmic musical creativity.

Participants	7 musicians
Jam protocol	jam session (12 minutes—all together) group interview (approx 30 minutes)
Data collected	touch logs (150K OSC packets, total log size 3.3MB) audio recording (direct stereo out from the server)

Table 4.3.: Jam session details for the Viscotheque v1 jam session.

4.3. Jamming with Viscotheque v1

On the evening of April 8, 2009 seven musicians were invited to come and participate in the Viscotheque v1 jam session. The musicians were all male, aged between 21 and 60. Each musician had at least one year of formal musical training, either in the classical or jazz tradition. The musicians were aware of the formal nature of the jam session. This was not a concert; there would be no audience. The musicians were all connected with the Australian National University (ANU) in some way, as students or lecturers in either the university's music school or computer science department.

The purpose of the jam session, from my perspective as the designer of the Viscotheque system, was threefold:

1. to examine the musical interaction which takes place as musicians use the instrument to jam;

2. to receive feedback on the design from jamming musicians; and

3. to test the system under the load of several musicians jamming simultaneously.

Although the jam session was held in the Computer Science building of the ANU rather than a smoky jazz club, care was taken with the sound and lighting to make the space as relaxed and suitable for jamming as possible. The jam took place in a small lecturing room which (when full) holds around fifty people, although there were only the seven musicians and myself in the room during the jam session. The 2.1 channel speaker setup was capable of providing satisfying volume—not 'rock concert' loud, but certainly a sound level which made conversation difficult without raising one's voice. The subwoofer in particular was useful for generating the low-frequency rumble of the larger drum sounds, such as the timpani. The lights were dimmed to provide a relaxing atmosphere, and the musicians were free to rearrange the seating in the room to their liking.

The musicians had never played the Viscotheque DMI before attending the jam session. Each musician brought his own iOS device, either an iPhone or an iPod Touch, and followed the four step installation and setup process described in the previous section to participate in the jam. I was on hand to assist them with any questions they had during the setup process.

[7] As an example, an interface where each touch of the screen directly triggered a drum hit would be an example of a 'direct' mapping.

Although I was present in the room throughout the whole session, I did not participate in the jam myself.

Once the musicians had all downloaded the MRMR app and connected to the server, the jam session began. No instruction about how the instrument worked was given beyond the labels on the interface widgets (see fig. 4.3(b)). No instruction was given on whether or not the musicians were allowed to talk to one another. There was some occasional discussion, particularly in the early part of the jam, but in the main the musicians were silently concentrating on their own iOS device (their instrument).

The jam session lasted for 12 minutes. Each movement of any of the control widgets was written with a time-stamp to a log file on the Viscotheque server. As well as the control data, the final (stereo) mix of the drumming ensemble was recorded.

After the jam was over, I conducted an informal feedback group interview of approximately 30 minutes with all the musicians. The musicians were encouraged to share their thoughts about the experience, as well as any suggestions for ways to improve the Viscotheque DMI. During the interview I took notes, but the interview was not recorded on tape.

From a technical standpoint, the jam session was a success. All seven musicians who attended were able to connect to the Viscotheque server with their 'BYO' iPhones and participate in the jam. The system handled this load without any glitches in the output audio, and no participants complained of unresponsiveness (such as would have been the case if packets were getting dropped).

Beyond the technical performance of the system, the Viscotheque v1 jam session was exploratory in nature, and there was no specific hypothesis about the interface being tested. Having built a DMI based on my own taste and experience as a jamming musician, the purpose of the jam session was to see how other musicians responded to the design, and to look for ways in which it could be improved.

4.4. Data visualisation

Each musician's 'voice' (at least from a sonic perspective) was wholly determined by their trajectory through the 5-dimensional input space of their drum pattern. Because of the low dimensionality of this input space[8] it is relatively simple to visualise these trajectories of interaction (fig. 4.5). Each parameter lay in the closed interval $[0, 1]$, and the slider was mapped linearly to this interval. The musical meaning of these parameter values—the effect they had on the output sound—is described in the mapping section (section 4.2). These traces reveal periods of high and low activity in the widget manipulations by the musicians. The high-activity periods are represented by solid 'blocks' of a single colour, in which the musician spent a period of time playing with a single parameter, 'zigzagging' the slider up and down through the full range of values. The drum type (purple) and beat period (blue) parameters seem to be the most prominent blocks of colour, particularly for musicians Mike and Josh.[9]

[8] In comparison to a more complex physical instrument, e.g. a guitar.

[9] As mentioned in chapter 1, the musician's names have been changed in this thesis to preserve their anonymity.

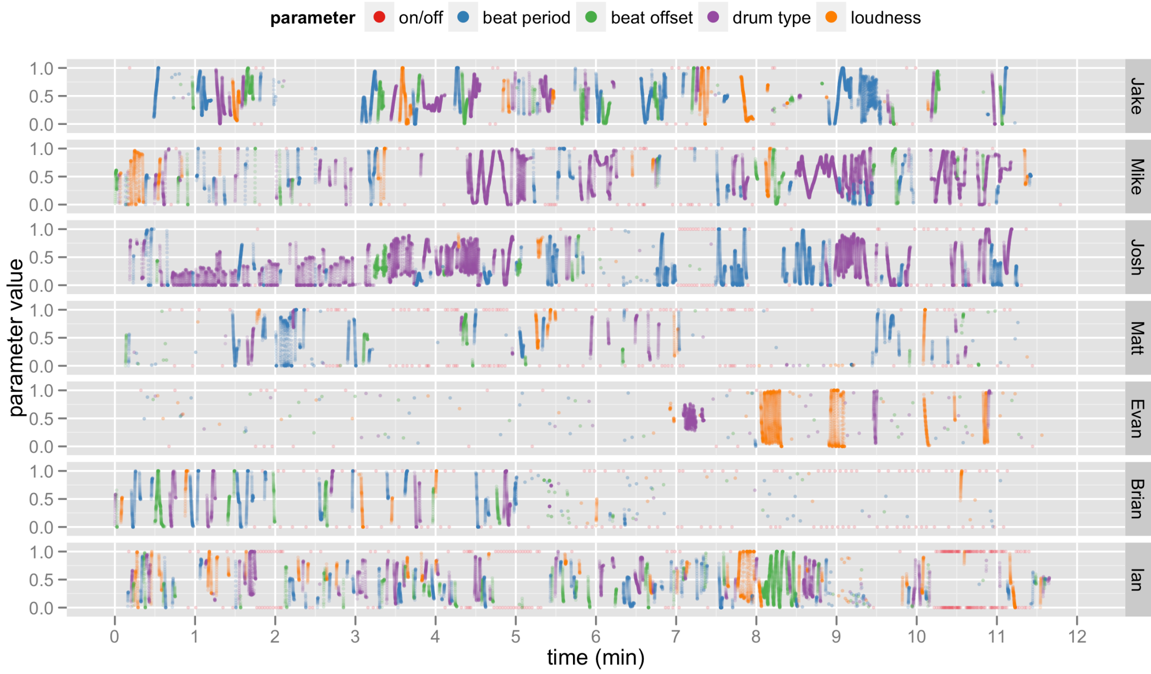
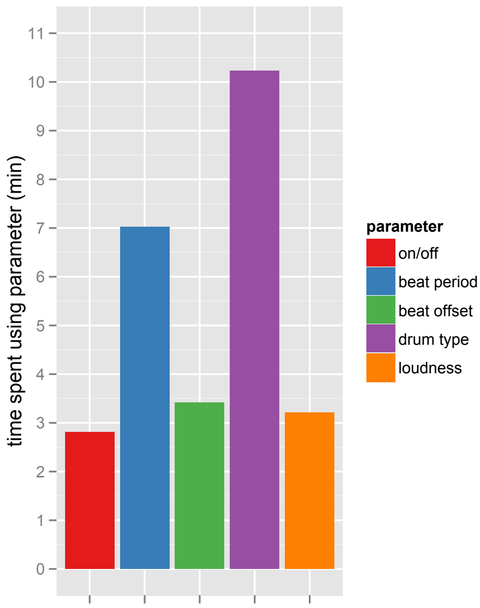


Figure 4.5.: Timeline of the drum parameter adjustments made by the musicians during the jam session.

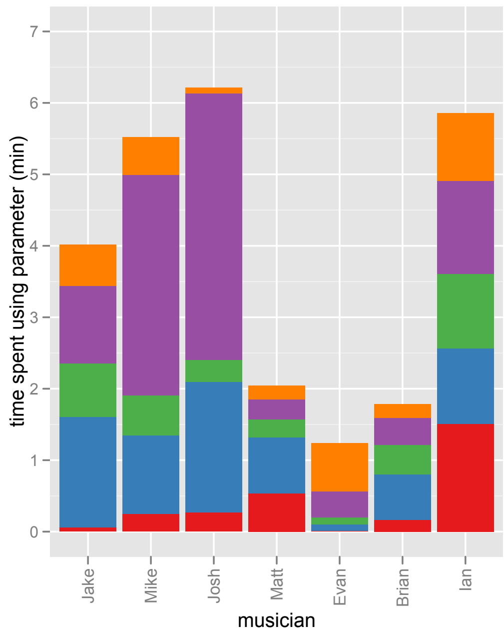
There are also some sparse regions in these touch traces, such as the first half of the jam session for Evan and the second half for Brian. These musicians were not inactive during these sparse periods—there are some parameter adjustments being made which show up as isolated 'spots' on the timeline. Some musicians exhibit both sparse and dense regions in the timeline. Interestingly, Evan starts out sparse and then changes tack to denser activity about half way through the jam—almost exactly the opposite of Brian, although this was most likely a coincidence.

In contrast to the timeline in fig. 4.5, fig. 4.6(c) is a view of the data aggregated over the whole jam session. As noted, fig. 4.6(a) shows that the 'drum type' parameter was the most manipulated, followed by 'beat period' (see table 4.2 for parameter descriptions). Looking at the breakdown by musician, it is clear that Mike and Josh are significant contributors to this trend, with both spending over three minutes (one quarter of the whole jam session) adjusting this parameter.

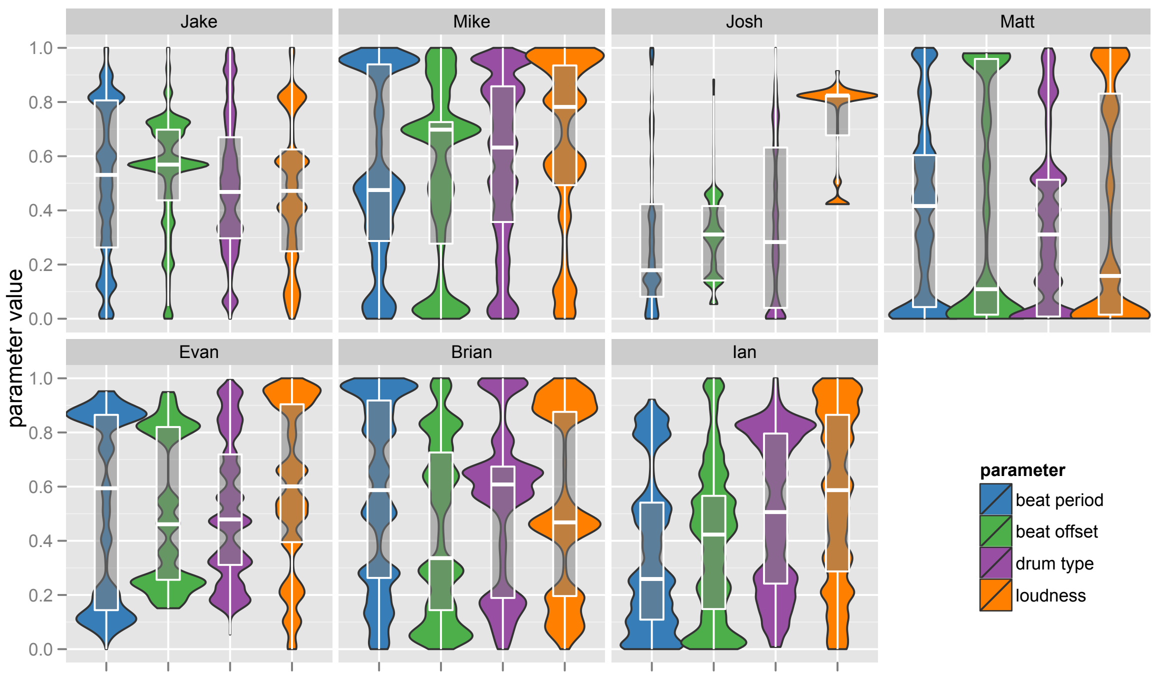
The drum type and loudness parameters are interesting in that they are the parameters which did *not* change the rhythm of the musician's drum pattern, unlike the beat period and offset parameters. When a drum pattern was playing, it was easy to hear the effect of changing the drum type or loudness because the timbre or volume would be affected while the rhythm would remain regular and consistent. The beat period and offset parameters, in contrast, would result in irregular rhythms (that is, irregular between-drumhit time periods) which would not settle down into a steady rhythm until the parameter adjustment



(a) The total amount of time spent manipulating (i.e. with a finger on the touchscreen) each parameter.



(b) The amount of time spent manipulating each parameter broken down by musician and parameter.



(c) Violin plots for the value of each parameter over the whole jam session. This includes the times when the parameter was not being actively adjusted by the musician—the parameters remained constant unless they were changed using the corresponding slider widget.

Figure 4.6.

was finished. This meant that it was more difficult for the musician to predict the effect that changing those parameters would have on their overall sound—a comment that came out in the discussion at the end (see section 4.5).

The violin plots[10] show the distribution of values for each parameter by musician (fig. 4.6(c)). In some cases, the distribution is strongly peaked around just a few (or even just one, such as Josh's 'loudness' value) values, while in other cases the density is fairly consistent over the whole parameter range. Again, this indicates a divergence of styles, from the 'set and forget' approach (**Matt**, **Evan** and **Brian**) to continuous use over the whole input range (**Jake**, **Mike**, **Josh** and **Ian**).

In terms of overall activity, there is a large disparity between the most and the least active musicians. Josh, Ian and Mike all had over five and half minutes worth of parameter manipulation, while Matt, Evan and Brian were all under two minutes. Jake is in between these groups, with four minutes of activity. This suggests that the *low activity* musicians played the instrument in discrete steps, changing the parameter values then letting the drum beat steadily away by itself, while the *high activity* musicians played the instrument by continuously varying the parameters.

It is important to recognise that this notion of *activity*—having a finger on the screen and manipulating a parameter—is not necessarily correlated with musical creativity, skill or engagement. The interface was designed as a process control interface specifically so that the musicians *could* make music without having to constantly interact with the device. What does seem to be true, though, is that different musicians employed different interaction styles at different times. In terms of navigating the input parameter space, some preferred discontinuous jumps while others favoured smooth trajectories.

Given the preliminary nature of this concert, I have not drawn any particular statistical conclusions from the interaction data. The purpose of looking at the data is to get a picture for how the musicians went about the practice of jamming on a mechanical level, and to set the scene for a discussion of the post-jam interview.

4.5. Musician interviews

The post-jam group interview was conducted immediately after the v1 jam session finished. As mentioned in section 4.3, during the interview I took notes based on the discussion. This required me to paraphrase the discussion in parts, and as a result there are not many direct quotes recorded from the discussion. This was not a problem for the Viscotheque v2 and v3 jam sessions, which were both recorded on video.

The interview began with the question *"so, how did you find it?"* and lasted for about 20 minutes. The question was deliberately open-ended. Because this was the first real outing of Viscotheque as a jamming DMI, any aspect of the musicians' experience was potentially helpful in refining the design of either the instrument itself or some other aspect of the

[10]Violin plots (Hintze, 1998) are a combination of a density plot and a boxplot, useful for displaying the distribution of a one-dimensional variable. The 'width' of the violin represents the distribution (as a density function) of the variable, so the coloured region is 'thicker' in the places where the variable occurred more frequently. The overlaid boxplot gives summary information about the median and quartiles of the data.

jam session. The ambiguity of the word *"it"* in the question leaves room for the musicians to interpret the question as being about the the physical (iOS) device, the touchscreen interface, the mapping, the volume level, the overall experience, or any combination of these.

After the initial question was asked in the interview no further specific prompting was required. The musicians were able to sustain the discussion with their thoughts about the experience, and they discussed the experience with each other as much as with the interviewer (me). Overall, the musicians enjoyed the jam session and were positive about the prospect about participating in future Viscotheque jam sessions.

To summarise the discussion, the issues which were raised can be grouped into four factors: two positive and two negative. The two factors which contributed positively to their experience of the jam session were the novelty and 'groove' of the DMI, while the two factors which contributed negatively were issues of legibility and expressiveness.

4.5.1. Novelty

Because they had never played it before, the musicians described the jam session as an opportunity to *explore the possibilities* of group music making with the instrument. This exploration was described positively, both from the perspective of it being a new avenue for creativity and also from the perspective of the interface being a 'puzzle' to figure out.

There was discussion in the post-jam interview about how exactly the mapping worked, with each musicians sharing their understanding of the influence of the different parameters on the output sound. There were some revelations, as musicians who had not understood the influence of a particular parameter were enlightened by those who had understood it. Overall, the musicians collectively showed a good understanding of the mapping, although certain individual musicians were unsure about certain aspects of the interface and sought explanation from each other during the interview. This is unsurprising given that they were figuring out the mapping on the fly, and indeed there was some discussion amongst the musicians even during the jam session about the way the instrument worked.

The musicians did describe some frustration associated with the novelty of the instrument, such as moments where they had a certain musical goal in mind (such as a particular rhythm) but were unable to make it happen due to their unfamiliarity of the instrument. A key cause of this frustration was a lack of legibility in their sound, as I shall discuss in section 4.5.3.

However, the novelty of the Viscotheque v1 instrument was held to be a positive influence overall on the musicians' experience in the jam session. The potential of a new instrument for music making offers a fresh opportunity to find a voice, and all of the musicians expressed a desire to jam again.

4.5.2. Groove

The musicians described moments where they were really enjoying the collective drumming of the group. This enjoyment is subtly different to the pleasures of exploration and 'figuring out' associated with the novelty of the instrument. These were periods where the music they were making was compelling in itself, and it was a satisfying feeling. The word used to describe the moments of musical satisfaction was *groove*. This term was used without prompting to describe the peak moments in the jam from a musical perspective. The use of this word in this way is unsurprising given the musical background and experience of the participants.

The moments of groove involved toe-tapping and head bobbing, both of which were visible at times as I watched the musicians during the jam session. The musicians praised the sound quality and volume level, reporting that at times they really felt a *groove* happening, and the satisfying sense of having the sound fill the room was cited as a influence in the perceived sense of groove. The subwoofer was helpful in this regard, as it provided a satisfying thump from the low pitched drums which would 'anchor' the rhythm, with the higher pitched drum sounds adding accents and colour.

One key characteristic of these moments of groove described by the musicians was their fragility. There *were* moments that worked, when everyone was drumming together, but did not always last:

Matt there were some moments where I was really enjoying the drumming and hoping it would continue, but then somebody changed something and that moment was gone.

This sentiment was echoed by many of the musicians—that sometimes they were feeling the groove, but then one of the *other* musicians would alter their sound, and the groove would fall apart.

This is in fact encouraging—peak moments in jamming is are the result of *collaborative emergence*, as discussed in section 3.1. In a jam session the overall outcome is always vulnerable to changes beyond the control of any one musician. The fact that even with the simple mapping used in Viscotheque v1 the musicians experienced this frustration is encouraging (even though it was frustrating for the musicians!).

The interview did not uncover any systematic pattern or antecedents for the emergence of a groove. Still, the fact that there were moments which the musicians described using the language of 'groove' are encouraging.

4.5.3. Legibility

The musicians did find some aspects of the instrument and jam session frustrating. Apart from the frustrations associated with not being in complete control and not working together properly discussed in the previous section, they also described some shortcomings in the instrument itself. At times, musicians found it hard to ascertain which drum they were controlling. This was a feeling shared by all the musicians.

The term *legibility* has been used to describe the extent to which participants in interactive environments

> know how their actions affected the performance development, and to understand the domain space of what actions they could take. (Taylor et al., 2011)

In the Viscotheque jam session, the musicians inability to determine which drum sound in the 'circle' was theirs is an example of poor legibility.

The musicians found some of the instrument's control parameters to be more legible. The 'drum type' and 'loudness' parameters were better in this regard because they only affected the *nature* of the drum hits, not their timing. This mean that the regularity and predictable timings of the drum hits could be used to keep a 'fix' on the particular sound under the musician's control even as those parameters were then varied. In contrast, when the beat offset and period parameters were changed the predictable timing of the drum hits was disrupted, and it was easier to lose track of one's sound in the mix (although the 'drum type' and 'loudness' parameters would stay constant in that case).

This preference for parameters which affect the *what* rather than the *when* of the musician's sonic output were reflected in fig. 4.6(a) and discussed in section 4.4.

Helpfully, the musicians suggested three ways in which this legibility situation could be improved:

1. individual (rather than shared) audio output streams;
2. visual feedback, perhaps some representation of the state of all the different musician's drum loops; and
3. more 'direct' mappings—controls which had a more immediate and predictable effect on the sound.

The pros and cons of using shared (rather than individual) loudspeakers was discussed in section 4.1, and the decision was made to go with a shared speaker setup. A visual feedback was planned to be incorporated in version 1 of the instrument, but it was not completed in time for the jam session. Visual feedback was part of Viscotheque v2 and v3. Similarly, a direct mapping was a core design goal of the later versions of the instrument.

It is also important to remember that the musicians had only jammed for *twelve minutes*, so to expect an intimate knowledge of the mapping and the effects of the different parameters after this short period is unreasonable, even for a deliberately simple 'process control' DMI with only a few parameters. Still, one of the opportunities of using smartphones in DMI design is to lower the barrier to participation in jamming, so the learning curve of the instrument is an important factor to consider.

4.5.4. Expressiveness

The final complaint that the musicians had with Viscotheque as a DMI for jamming was to do with its expressiveness. By this, I mean the amount of variation possible in the output sound, or the degree of freedom in music making. Some of the musicians felt that the simple, abstract interface of five parameters may not provide sufficient interest in the long term. It is not that they were bored by the end of the jam session; rather that they felt that they would run out of creative possibilities sooner than they would like.

Other musicians disagreed with this point, feeling that the constrained interaction space could lead to more reflective music making. The rationale was that this simplicity freed them from the worries of having to operate all the nuts and bolts of the music making process and allowed them to concentrate on higher-level aesthetic functions, such as the overall dynamics of the jam session.

In practical terms, the way to offer the musicians a more expressive instrument is to offer them more control dimensions along which to vary their sound. The easiest way to do this is to simply increase the number of parameters which control the output sound. There are many audio parameters which the v1 system does not expose to the musician, particularly if the sound output expands to include pitched instruments as well as unpitched drum sounds. Although there was no consensus about the right number of parameters to expose in a DMI for jamming, this was an important reminder that this aspect of the interface (the input dimensionality) was worth reconsidering for the next version of the interface.

4.6. Chapter summary

This chapter presented version 1 of the Viscotheque instrument; the architecture, the design process (especially the gesture→sound mapping) and the results of evaluating the instrument in a jam session with real musicians.

This first phase of the Viscotheque experiment was an exploratory one. It showed that there *was* promise in using simple, smartphone based music making tools in a jamming context. The musicians in the jam session described glimpses of the satisfying groove experience of jamming (section 3.2), and identified potential shortcomings of the instrument in the areas of legibility and expressiveness. These factors were useful as a starting point for the second iteration of the system (Viscotheque v2) which is the subject of the next chapter.

5. Viscotheque v2

The development of version 2 of Viscotheque took place between April 2009 and February 2010. Incorporating the lessons learned from Viscotheque v1, the goal with version 2 was to further explore low-dimensional 'process control' interfaces for jamming. In v2 a more expressive interface was given to the musicians, allowing them to control not only *when* the samples were triggered but also to 'morph' them on-the-fly. Viscotheque v2 also included a visual feedback element not present in v1.

Three jam sessions were held to evaluate v2 of the Viscotheque DMI, each with different musicians. The post-jam interviews were conducted using a video-cued recall technique (see Costello, Muller, et al., 2005) to elicit feedback about the salient moments of the jam session, both positive and negative.

The work presented in this chapter was presented in the long paper B. Swift, H. J. Gardner, and A. Riddell (2010). "Engagement Networks in Social Music-making". In: *OZCHI '10: Proceedings of the Australasian Conference on Computer-Human Interaction.*

5.1. Architecture

Viscotheque v2 used a client-server architecture with the same 'thin client + sound processing server' structure as in v1 (see section 4.1 for details). The two key changes made between v1 and v2 were the addition of visual feedback for the musicians and the switch from the MRMR iPhone app (*MRMR - Open Mobile Touch Protocol*) to a custom Viscotheque app developed in-house.

5.1.1. Viscotheque server

The 2.1 speaker set-up and jam room were kept the same as in v1. The visual feedback required extra hardware in the form of a data projector and projector screen. The visuals were generated in real-time by the Viscotheque server Impromptu application. As mentioned in section 4.1, the server software environment Impromptu is designed for graphics as well as audio, and was well suited to this task. The exact nature of the visual feedback is discussed in section 5.2.

The Viscotheque server application was significantly rewritten to provide the different sound mapping used in Viscotheque v2, although the server was still written in Impromptu as before.[1] In version 1, the audio processing required to provide the Viscotheque mapping was straightforward—triggering the playback of a small set of drum samples. In Viscotheque v2 the audio processing was much more complex, including real-time pitch shifting and filtering. Impromptu operates as an AudioUnit host, so the audio processing was offloaded

[1] The server application was 228 Source Lines of Code (SLOC) in scheme.

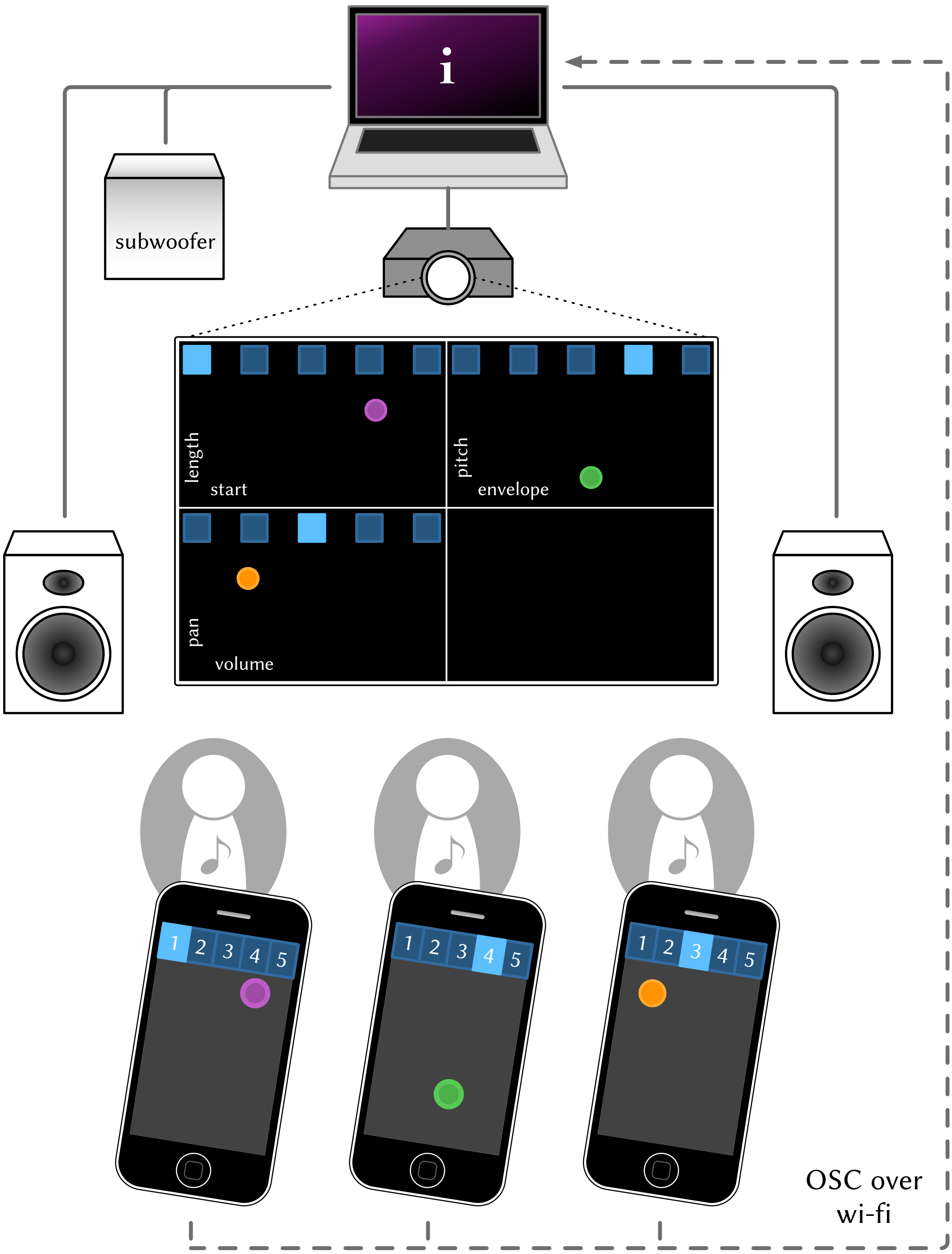


Figure 5.1.: Viscotheque v2 system architecture. The three musicians were co-located and seated in a semi-circular configuration facing the screen. The screen showed real-time video feedback about the current mode and finger position of each musician. As in v1, all sound output came through a pair of stereo speakers in the jam room. The iOS devices communicated with the Viscotheque server via OSC over the WLAN.

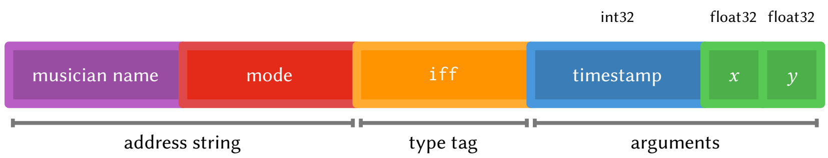


Figure 5.2.: Viscotheque v2 OSC control message structure. The mode portion of the address string indicated which mode the interface was in (see table 5.1 and fig. 5.4(b)).

to Native Instruments' *Kontakt 3* sampler. It was therefore not necessary to write the audio Digital Signal Processing (DSP) code in Impromptu (although Impromptu is capable of sample-level audio manipulation as well).

The client iOS application sent lightweight control messages (just an indication of the x and y position of the current touch–see fig. 5.2) to the server. The Impromptu server's role was to parse the control message from the device, trigger the appropriate processing and playback (through Kontakt), and generate the visual feedback shown on the screen. This was all done in real-time, with a goal of minimal latency so as to provide a responsive interface to the musician playing the instrument.

5.1.2. Viscotheque client

The v2 iOS client application was an iOS application that I designed and built in-house for Viscotheque.[2] This was the most significant change to the Viscotheque infrastructure from v1 to v2. I chose to do this (rather than continue to use the MRMR app) firstly because it gave me complete control over the way the app worked and would continue to work in the future. The iOS App Store[3] works as a centralised distribution channel for all iOS software—it is the only way to get applications on an iOS device.[4] Even though the MRMR app was open-source, its developers had control over its presence in the app store and the timing of its updates. If the functionality of the app was changed by the MRMR developers then there was no way to provide an older version or patched version of the application to the jamming musicians. This had not caused any specific problems in the version 1 jam session, but had caused a problem in a related project. The risk of not being in control of the application nor having an easy way to provide an alternative if things were changed was key reason for developing a new Viscotheque application.

The other reason that the app was rewritten was to provide more control over the interface. The MRMR interface protocol[5] (which was used to specify the widgets which would appear on the screen) was limited to a collection of widgets (sliders, buttons, touch zones) specified by the MRMR app developers. Custom-coloured buttons, for instance, were not available.

[2] The Viscotheque iOS application was 1450 SLOC in Objective-C.

[3] `http://www.apple.com/iphone/from-the-app-store/`

[4] This is not *strictly* true, there are other ways with developer accounts or jail-broken devices, but from the perspective of the average user the App Store is the only distribution channel.

[5] The MRMR interface protocol is documented at `http://mrmr.noisepages.com/mrmr-interface-protocol/`

By writing a custom iOS app, any interface could be presented to the musicians, which was particularly useful for using different colours to differentiate between musicians on the main visual feedback screen in the jam room.

The Viscotheque v2 application required very little setup before the musician could start jamming. The manual IP address configuration step which was part of using the MRMR app in Viscotheque v1 was no longer necessary due to the use of zeroconf[6] (Steinberg and Cheshire, 2005) 'zero configuration networking' to automate the device discovery and IP address setup process between the clients and the server. This process worked by having the Viscotheque server advertise on the local network that it provided the service "Viscotheque". When the app was launched it presented a welcome screen (fig. 5.4(a)) to the musician, but also searched the network for the "Viscotheque" service. If the server was running, the device would receive in response to the query the IP address and port number on which the server was listening for OSC packets. This process was automatic and required no input from the musician.

So, to start jamming with version 2 of the Viscotheque DMI required only *four* steps, down from five in version 1:

1. Download the Viscotheque app from the App Store
2. Join the Viscotheque WLAN
3. Launch the Viscotheque app
4. Start 'playing' the instrument using the touch interface on the device's screen

5.2. Mapping

The Viscotheque system allowed the participants to simultaneously influence the playback of multiple audio loops. Building on the drum circle idea of v1, Viscotheque v2 gave the musicians control over longer segments of audio. While in v1 the musicians could only control *when* their audio segments were triggered, v2 also allowed them to 'morph' the audio in real time. Each musician was given a different audio file as 'source material', but they could control which segment of that audio sample they were looping. This gave the musicians some variation in the audio they were looping—it was always be a subset of the audio file they were given, but they had control over and could change which subset was being looped at any given time.

To facilitate this, more control parameters were exposed to the musician. The loop control parameters of v1 (beat period, beat offset) were present in v2. In addition, the 'loop start' and 'loop length' afforded each musician fine grained control over which segment of audio is currently looping.[7] Based on these four parameters, each musician produced a repetitive

[6] also known as Bonjour

[7] the term 'loop' is (perhaps confusingly) used as both a noun and a verb in the context of sound editing and production. As a noun, the loop is the *unique segment of audio data* that is being played back, and the *process* of repeatedly playing this audio data (starting again from the beginning when the end is reached) is called looping.

Mode		Parameter	Description
1	x	loop start	position of current selection in audio file
	y	loop length	length of selection
2	x	loop offset	deviation of playback from 'downbeat'
	y	loop period	frequency at which section is looped
3	x	pan	playback stereo position
	y	volume	playback volume
4	x	envelope	attack/decay envelopes for playback
	y	pitch	playback pitch
5	x	cutoff	cutoff frequency for lowpass filter
	y	resonance	filter gain for lowpass filter

Table 5.1.: The control mapping for the Viscotheque v2 instrument.

stream of sonic material. A visual representation of this process is shown in fig. 5.3.

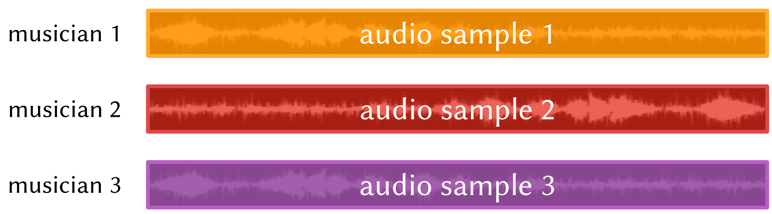
The audio files were all instrumental guitar patterns lasting around 30 seconds. This audio material was chosen because the guitar provides a good balance of rhythmic and melodic/harmonic content. This material was not just for playing back unprocessed, however. Volume, pan, and several timbral manipulation parameters were exposed to the musician as well. While the loop was 'looping', these parameters could be used by the musician to creatively morph the output sound. This was a change from the v1 interface, where the musicians had control over *when* their drum hits were played, but little control over *how* they were played—the samples were played back largely unprocessed. In v2 of the Viscotheque DMI the musicians could control the filter and resonance parameters of a lowpass filter, pitch-shift the output in real-time, and control the playback envelope. This allowed them more dimensions to explore and more scope for creative variation. A complete list of the ten parameters available for manipulation in Viscotheque v2 is provided in table 5.1.

As a control device, the input and output dimensionality of Viscotheque v2 was greater than v1 (as described in section 4.2).

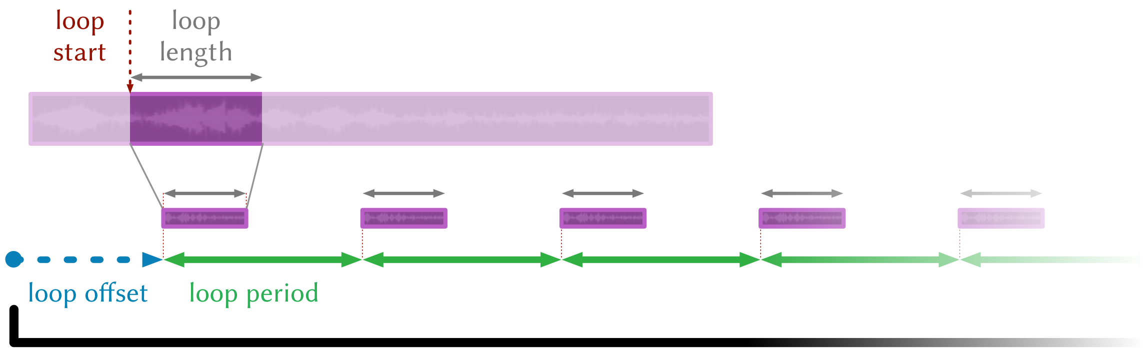
Input space
There were ten control dimensions, all of which were continuous. Mathematically, this is the set $x \in [0, 1]^{10}$. Unlike v1, there was no on/off button—although the volume parameter could be set to zero to silence the instrument.

Output space
The same as in v1: the space of all possible digital bitstreams. In practice, though, only a vanishingly small fraction of these will be output by the instrument.

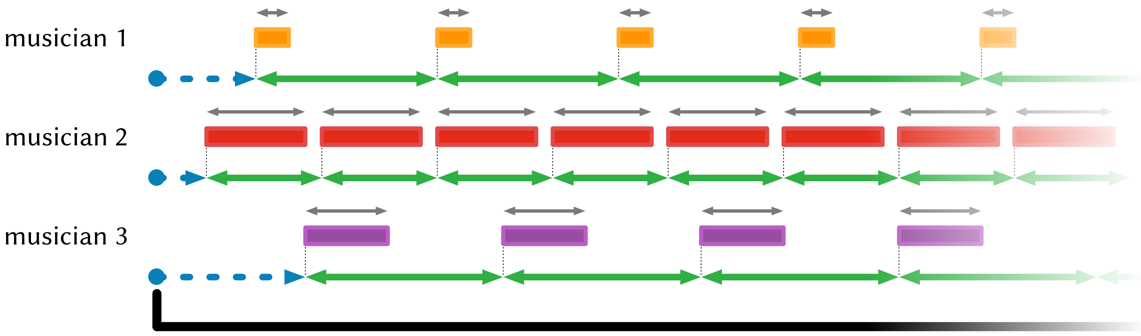
Both the input and output dimensionality of the Viscotheque instrument were greater in v2 than in v1. The latter is a consequence of the former—more parameters to control means a greater range of ways to affect the sound, and a greater range of sounds and expressive



(a) Each musician was assigned a different audio sample; this was their 'raw material' to manipulate.



(b) The musician selected a portion of their audio sample to loop by adjusting the loop start and loop length parameters. This loop was then played back with an given offset and period as per Viscotheque v1.



(c) An example of three musicians jamming together. Each of the parameters could be modified in real time.

Figure 5.3.: In Viscotheque v2, rather than controlling only a 'duration-less' drum hit (as in v1), each musician controlled the looping and manipulation of a segment of audio. As well as controlling the repeat (temporal) pattern of their sound, the musicians could pitch-shift or filter their sound, a well as control the volume and stereo panning. This afforded them a broader palette for sonic experimentation.

options were afforded the musician. This was to address the limited *expressivity* of the v1 interface, an issue raised by the musicians in the v1 jam session.

The v2 interface for controlling the ten parameters (see table 5.1) was qualitatively different from the interface in v1. In v1, the five parameters were controlled by five distinct widgets. However, this one-to-one relationship between parameters and widgets becomes a problem as the number of manipulable parameters increases. Five (usably-sized) sliders fit on the iPhone's screen, but this is approaching the limit of usable widget density. For v2, then, a modal interface was used, with two-dimensional 'touch zones' replacing the one-dimensional sliders the v1 interface. Instead of having all parameters available at once, five different 'modes' each provided control of pair of playback parameters, one mapped to x and one to y. The top part of the interface was composed of 5 buttons for switching between the modes, while the rest of the touchscreen was a 2D control pad. The interface is shown in fig. 5.4(b).

The MRMR application used in Viscotheque v1 was not able to provide the interface shown in fig. 5.4(b), which was a key reason for developing the iOS app in-house. As in v1, the iOS app was still a thin client sending control messages, sending in each OSC message a representation of the x and y position of the current touch plus the current 'mode' the musician is in (see fig. 5.2). The server would receive these control messages, apply the appropriate processing and adjust the sound playback accordingly.

The final part of the Viscotheque v2 instrument was the visual feedback. This was added in response to comments from the musicians who participated in the v1 jam session. The visual feedback was generated by the Viscotheque server and displayed on a (3 metre diagonal) projection screen in the jamming room. Because the control messages from each device contain all the information about the musician's current state (current mode and finger position), generating real time visuals was simply a matter of utilising Impromptu's Quartz 2D bindings to generate the appropriate feedback.

The display (which is shown in fig. 5.1) was a one-to-one representation of each musician's iPhone interface and current touch position. The screen was segmented so that each participant's screen was represented in its own segment, and each participant's touch point indicator was colour-coded to match the colour on their own device screens.

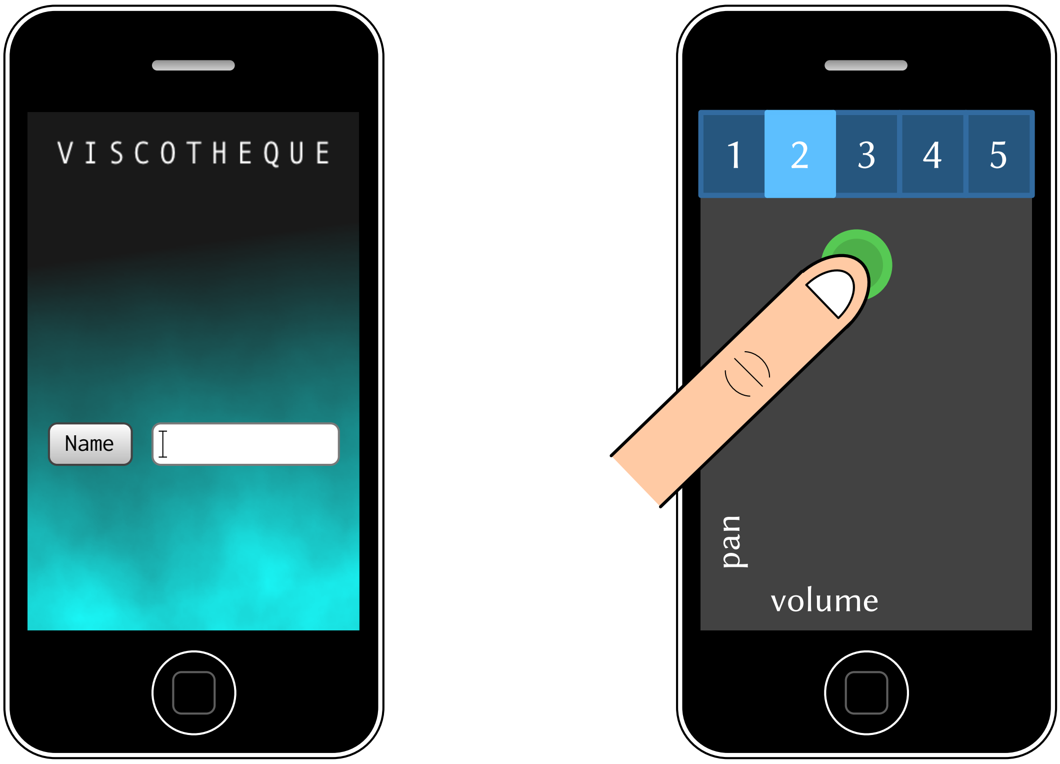
The design goals of the visual feedback were twofold:

1. to aid the musicians in orienting themselves and figuring out which sound in the overall mix was theirs (the problem of *legibility*, see section 4.5.3)

2. to allow the musicians to see what their collaborators were doing—what mode they were in and what gestures they were performing

5.3. Jamming with Viscotheque v2

Once version 2 of the Viscotheque DMI was finished it was given to musicians to play around with in a jam session. Unlike the v1 jam session, where all seven participating musicians jammed together in one large group, in v2 multiple smaller jam sessions were held, each with different musicians. There were three jam sessions, each with three musicians, conducted on the 26th and 27th of February 2010. These jam sessions were conducted in the same jam



(a) The Viscotheque v2 welcome screen. While this screen was displayed, the application was searching for (and advertising its presence to) the Viscotheque server. Upon receiving a response, the interface would change to the music making interface shown in fig. 5.4(b).

(b) Viscotheque v2 music making interface. The majority of the screen was a two-dimensional touch zone (control surface) for manipulating the musician's audio loop. The five buttons at the top of the screen were for switching between the different modes (the current mode's button was highlighted in a different colour). In each mode, a different pair of audio manipulation parameters were mapped to the x and y axes of the touch zone, and the current parameters under control were displayed in the bottom left-hand corner of the touch zone.

Figure 5.4.: Viscotheque app version 2 home screen and music making interface.

Participants	9 musicians, in 3 groups of 3
Jam protocol	solo practice (3 minutes each—musicians take turns) jam session (16 minutes—all together) group VCR interview (approx. 40 minutes)
Data collected	touch logs (215K OSC packets, total log size 14MB) audio recording (direct stereo out from the server) video of jam session & interview interview transcripts (9340 words total)

Table 5.2.: Jam session overview for the Viscotheque v2 jam session.

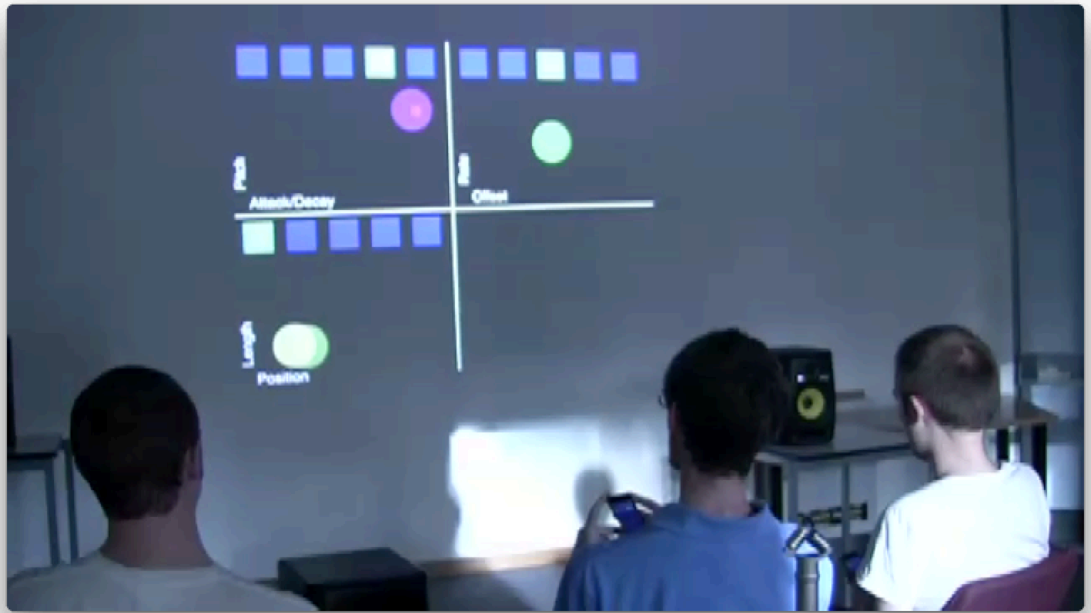


Figure 5.5.: Three musicians during a Viscotheque v2 jam session.

Group 1	Group 2	Group 3
Sean	Joe	Ryan
Andy	Dan	Ted
Zoe	Will	Alan

Table 5.3.: Version 2 jam session participants. These are not their real names, but the gender of each musician has been preserved in anonymisation.

room as the v1 jam sessions (see section 4.3). Again, the musicians were co-located, seated so that they could see each other as well as the shared screen (see fig. 5.1).

The reason for having multiple smaller jams rather than one big one was due to the increased expressivity of the v2 interface. As discussed in section 4.5, the musicians did find it hard at times to discern their own sound from that of the other musicians. Because each musician now had a greater range of sonic possibilities due to a more expressive interface (as described in section 5.2), there was an even greater risk that the musicians would get lost in this way, so having smaller groups (along with the visual feedback) was an attempt to mitigate this risk. Having multiple jam sessions with different musicians also presented the opportunity to compare the different jamming approaches of the different groups.

Nine musicians (eight males and one female, aged 23 to 30) participated in the v2 jam sessions, divided into three jam sessions with three musicians in each. The division of the musicians into groups is given in table 5.3. Like the v1 jam session, all participants had at least one year of musical training in either the classical or jazz tradition, and were recruited from the ANU's music school and Canberra music community. None of the participants in this round had participated in the v1 jam session, and the groups did not know each other beforehand. As noted by (Blaine and Perkis, 2000), musicians with musical expertise have experience in situations where they are required to listen to an audio stream and identify their own unique part of that audio stream—discerning their own effects from the effects of the others. This is an important skill in jamming with the Viscotheque instrument (see the v1 post-jam interviews in section 4.5). While it would be interesting to compare the difference in interaction patterns between musicians and non-musicians, this was not the purpose of the Viscotheque jam sessions.

At the beginning of each jam each musician (in turn) was given a three minute solo period in which to familiarise themselves with the sounds the instrument could make. During this period, the other musicians' iPhones were inactive. This 'practice session' was a suggestion from the musicians after the v1 jam session. After the practice sessions, the musicians were allowed to make music *together*. The jam session lasted for 16 minutes, making $3 \times 3 + 16 = 25$ minutes of jamming all up for each group.

For each jam session, the musicians were recorded on video (fig. 5.5 shows a still from one of these videos). The musicians were aware of the camera and the fact that the session would be recorded. Again, all the control messages sent to the Viscotheque server were logged, as well as a direct capture of the audio and visual data generated by the musicians in their jam.

The video was not captured solely for the purpose of archiving or analysis at a later date.

Immediately following the performance phase, I interviewed the musicians as a group while we together watched a video of the jam session. This Video-Cued Recall (VCR) technique has been used elsewhere (Costello, Muller, et al., 2005) to aid participants in describing their experience with interactive multimedia. The VCR interviews themselves were also recorded on video, and were later used to produce complete transcripts of the interviews.

As in the v1 post-jam interview, discussion between the musicians was encouraged. Asking specific questions of participants (as in a questionnaire) can be less useful than allowing them to describe their experience in a more open-ended fashion (Boehner et al., 2005).

The purpose of the VCR technique was to provide the musicians with stimulus material for their discussion, and also to allow them to point out any moments which were particularly memorable and describe the subjective experience of those moments. The VCR interview was conducted as a group, and so the musicians were also able to point out periods or patterns of interaction from other musicians and ask them about their recollections of those moments. As issues or vocabulary arose organically, I was able to ask follow-up questions to investigate these moments in greater detail.

5.4. Data visualisation

As was the case in examining the log data from the v1 jam sessions, it is easy to visualise the raw interaction data, but difficult to discern the meaning of this data in regard to the high-level compositional intentions of each jamming musician. I shall therefore restrict this analysis to considering patterns of touchscreen interface use by the musicians. The VCR interviews in section 5.5 are a better opportunity to understand the subjective experience of jamming and the musical intentions underlying the different moments in the Viscotheque jam session.

Figure 5.6 shows the parameter manipulation trajectories for all musicians in the jam sessions. The top three rows of the figure (which represent group 1) seem more ‘spread out’ over the whole parameter range than the other rows. In contrast, the other groups (particularly Joe from group 2 and Ted from group 3) show periods of solid colour around the centre line (representing the centre of the iPhone touchscreen and a parameter value of 0.5). This is most noticeable for the blue, purple and orange colours (modes 2, 4 and 5 respectively). This suggests an interesting statistic: the *mode time*—the time spent in a given mode before switching to the next one. Smaller values of this statistic correspond to quick changes between the modes, while larger values correspond to longer periods spent playing in a given mode. In fig. 5.6, the mode time is the length of a contiguous block of a single (pair of) colours.

In table 5.4, the ten longest mode times are shown. As suggested by the parameter traces in fig. 5.6, Joe (group 2), Ryan and Ted (group 3) dominate, taking up nine of the top ten places. The longest time spent in a single mode was 97 seconds by Ryan in mode 4 (envelope and pitch manipulation), approximately fourteen times longer than the median mode time of 7.3 seconds. The modes which make up these top ten positions were modes 2, 4 and 5. As described in section 5.2 these modes represent parameters which have a direct (volume or timbral) effect on the sound, as opposed to modes 1 and 3 which control the looping pattern.

musician	group	mode	time (s)
Ryan	3	4	97.4
Joe	2	5	85.8
Joe	2	2	78.7
Ryan	3	5	54.4
Joe	2	4	52.8
Ryan	3	5	51.6
Ryan	3	4	50.7
Will	2	2	48.0
Ted	3	5	46.9
Ted	3	4	46.7

Table 5.4.: Top 10 longest mode times (time between mode switches). The median mode time was 7.3s and the median absolute deviation was 7.6s. The distribution of the mode time statistic is shown in figs. 5.7(c) and 5.7(d).

This seems to be consistent with the v1 jam sessions, where the more direct parameters were also the most popular, i.e. the most used by the musicians.

Looking at the overall distribution of the mode time statistic in fig. 5.7(c), modes 2, 4 and 5 exhibit long tails, but the median values are similar. This indicates that although there were some long periods where musicians focused and stayed in a particular mode, in general the musicians were quick to change between modes in their playing. The breakdown of mode times by musician in fig. 5.7(d) does indicate that some musicians were quicker to jump between modes than others—again, musicians Joe and Ryan exhibit long tails (which is why they are represented so heavily in table 5.4), but Will actually has the highest median value. Group 2 is the most diverse with respect to these distributions, while in groups 2 and 3 the musicians are more similar.

Looking at the total time spent in each mode, there is not as large a variation between the different parameters as was the case in v1. In v1, the highest value was approximately three and a half times the lowest value. In contrast, the ratio of most-used to least-used parameter in the v2 jam sessions was only 1.4. This indicates that the musicians are showing less of a specific preference for any of the sonic control dimensions at their disposal. Similarly, there is a smaller difference between the most active and least active musicians (only twice as active vs six times more active in v1). This shows that the musicians may actually be showing less diversity in their interaction with the interface, or at least that the use or non-use of the interface (that is, having or not having a finger on the screen) is less notable as a differentiating factor between the musicians.

From a design perspective it is interesting that the increased expressivity of the v2 interface (in terms of the number of parameters under the control of the musicians) actually led to a *more* equal distribution of activity in each of these parametric dimensions compared to v1. Perhaps because the musicians had more options it took longer for them to explore the boundaries of the Viscotheque instrument, leaving them less time to ‘specialise’.

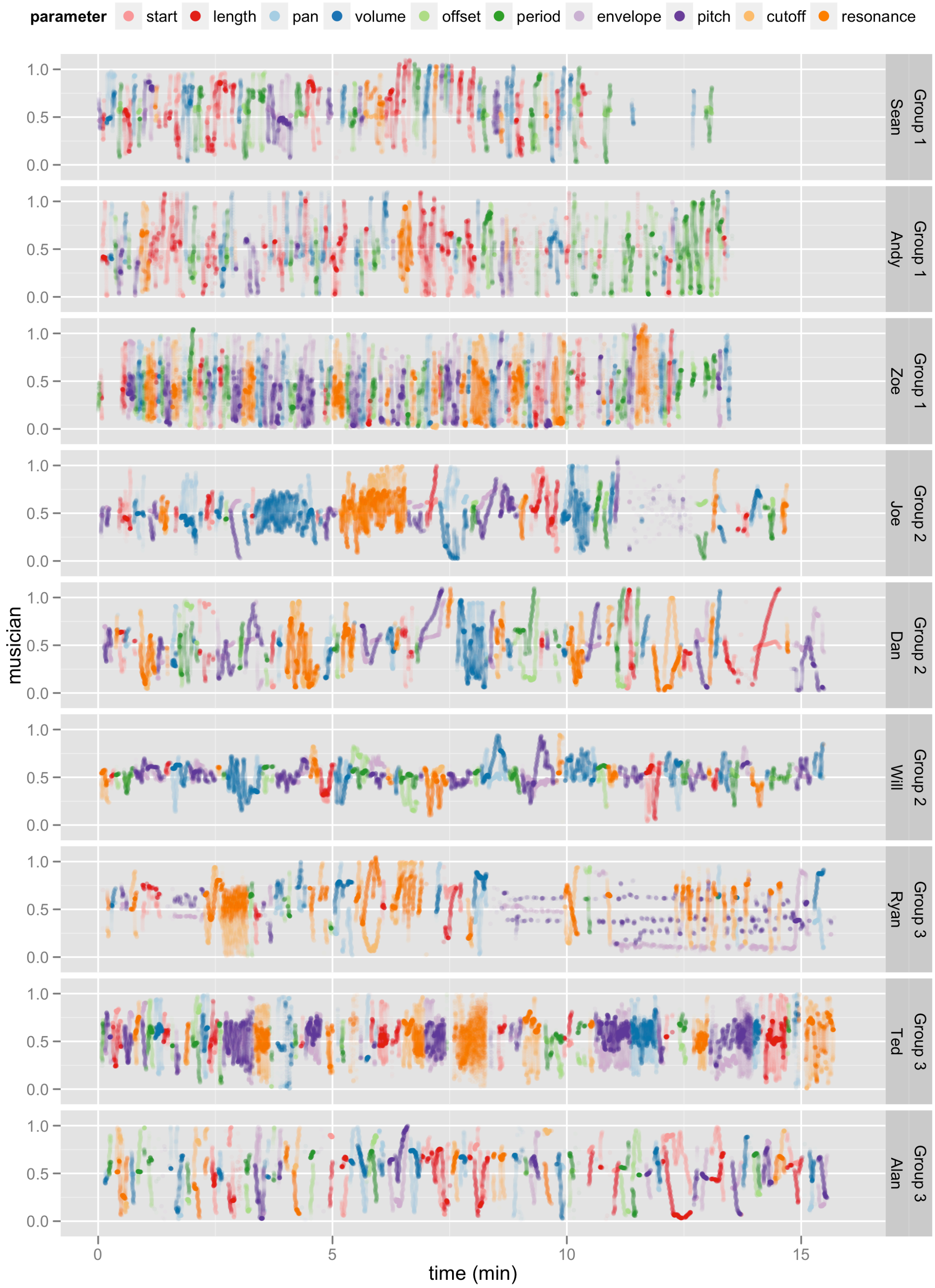
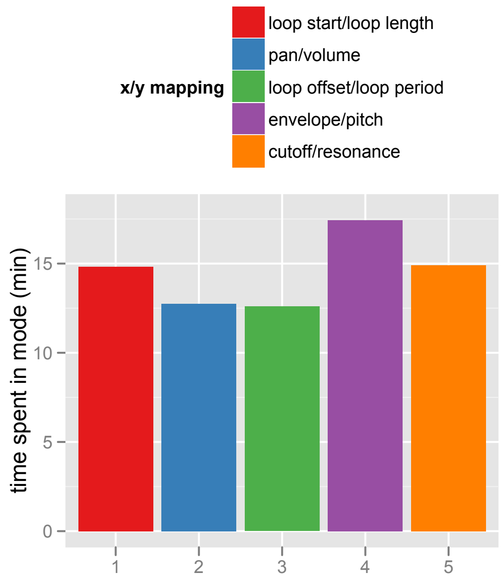
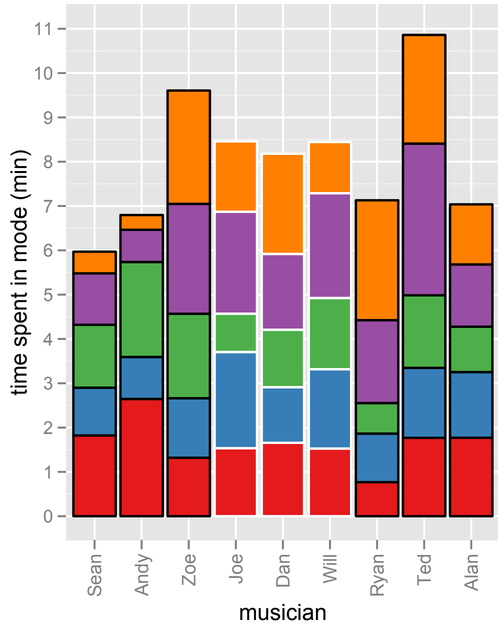


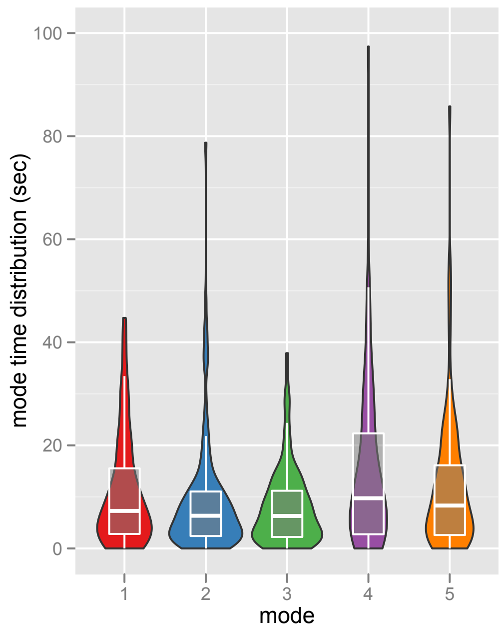
Figure 5.6.: Timeline of the parameter adjustments made by the musicians during the jam session. The musical meaning of each parameter is described in table 5.1 and fig. 5.3. Each finger touch actually changes *two* of the parameters (based on the x and y location of the touch), and these pairs of parameters are plotted with similar colours.



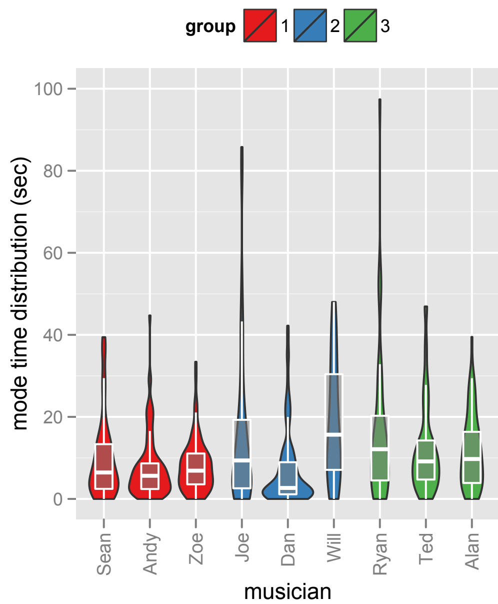
(a) The total amount of time spent in each mode. Note that in each mode the musician can control *two* parameters—mapped to the x and y touch position. The x/y mapping for each mode is shown in the figure legend (see table 5.1 for more detail).



(b) The amount of time spent in each mode broken down by musician and parameter. The outline colour (white/black) is used to group the musicians who were in the same jam session. The colours represent the same modes as in fig. 5.7(a).



(c) The 'mode time' distribution for each mode. This is the distribution of the time between mode switches.



(d) The 'mode time' distribution for each musician.

Figure 5.7.: Breakdown of the musician's activity by mode for the Viscotheque v2 jam sessions.

As well as the similarities and differences between the musicians in the way they switch between modes (and therefore switch between the parameters they are controlling), it is also informative to look at the values these parameters took over the course of their jamming. The parameters were mapped to the x and y position of a finger touch on the iOS device's touchscreen. Plots of these finger positions over the whole jam session are shown in fig. 5.8. Similarly, violin plots for the distributions of the parameters are shown in fig. 5.9.

The most obvious aspect of the finger position plots in figs. 5.8(a) and 5.8(b) is that there are noticeable differences between groups and between musicians, especially in regard to screen 'coverage'. Group 1 explored all areas of the touchscreen, while groups 2 and 3 left certain parts of the screen unexplored, even after 25 minutes of jamming. Will (from group 2) was the most conservative in his touch interaction, restricting himself to a cross-shaped area in the middle of the screen. This suggests a methodical and systematic approach to playing the instrument, as distinct from, say Zoe or Andy (from group 1) who seem to have been much more adventurous (at least from the perspective of screen coverage) in their touch traces.

Interestingly, when the parameter values are aggregated over all musicians across all groups for the whole jam sessions, the distributions are remarkably similar for each parameter (see fig. 5.9(a)). Each distribution is relatively symmetric and centred around the middle of its range (0.5). While some of the distributions exhibit heavier tails than others (representing more activity around the extremes of the parameter range), the overall effect is that they appear very similar. This is not the case, however, when these distribution are broken down by group. In many cases, the distributions of a given parameter are not peaked in the middle at all, and are in fact relatively uniform. It is interesting that this diversity does not carry through to the overall distribution plots, suggesting that these differences are group-specific and not to do with any particular inclination or bias of the interface.

So, to summarise the salient features of the data as shown by this visualisation process:

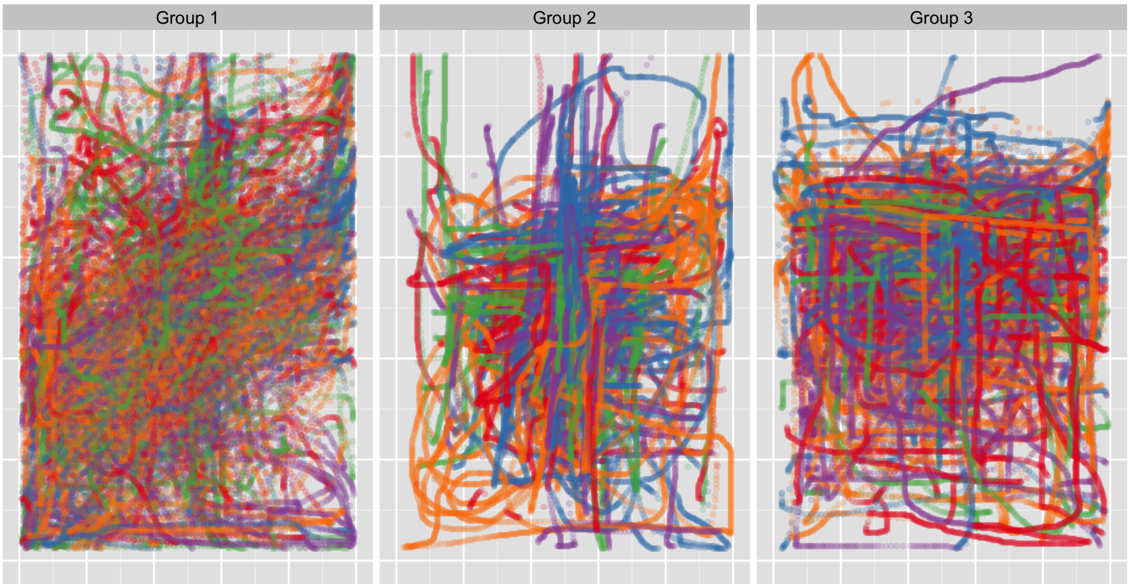
1. The musicians primarily switched regularly between modes, with a few longer segments (especially with parameters which had a direct perceptible effect on the timbre of the sound).

2. There was less variation in activity between the different parameters and between different musicians than in the v1 jam sessions. This may have been due to the shared visual feedback providing hints about 'styles of playing' between the musicians, leading to more coherence.

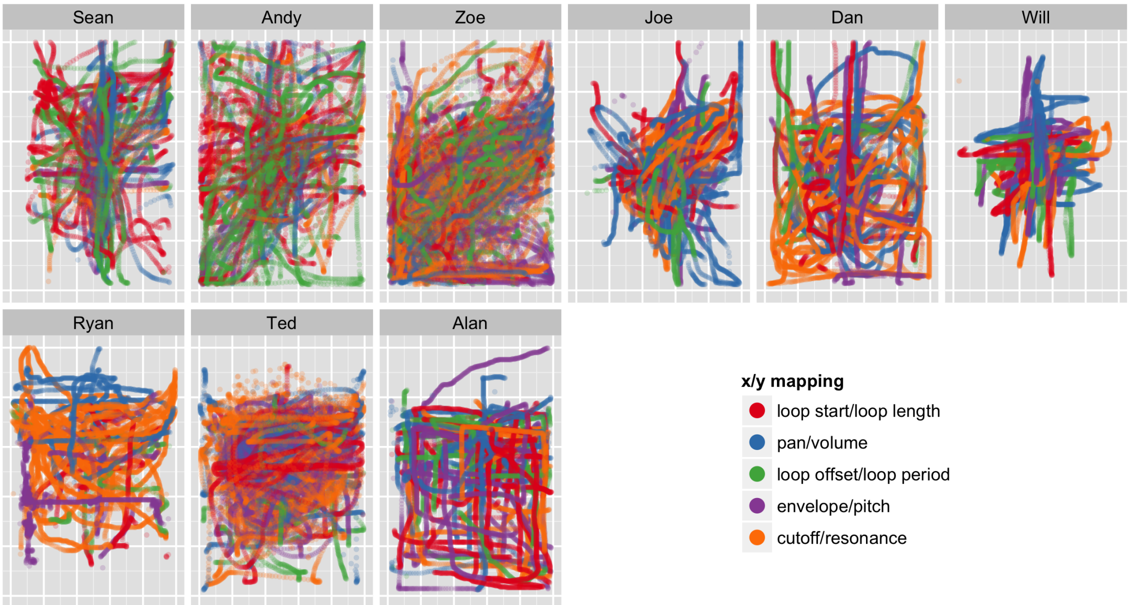
3. There were noticeable differences in touchscreen activity between the groups, especially in their 'coverage' of the touch zone.

4. There were differences between the overall distribution of each parameter when considered on a group-by-group basis, but when examined over all groups the parameters were distributed quite similarly.

As I mentioned at the beginning of this section, it is difficult to take any of these features and draw specific conclusions about the musicians' experience during the jam session. The post-jam interviews present a chance to explore these issues.

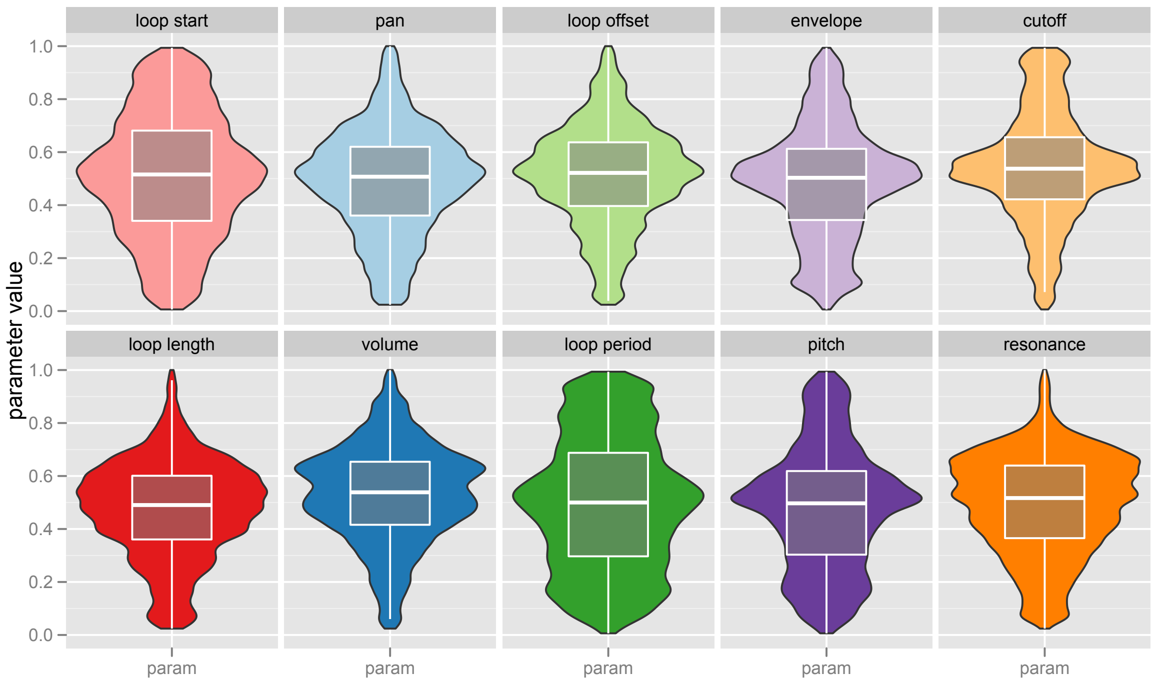


(a) Finger touch traces for each group of musicians. Each plot area is a one-to-one representation of the iPhone's 2:3 aspect ratio screen. A guide to which colours represent which interface modes is shown in fig. 5.8(b).

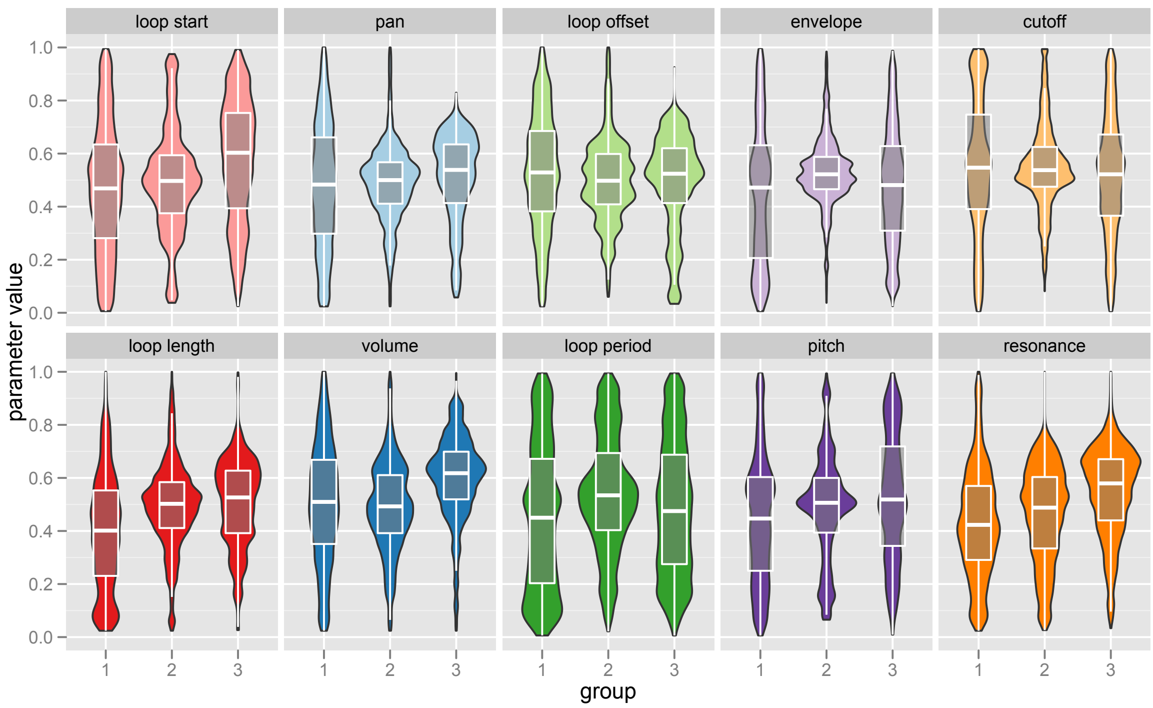


(b) Musician touch activity. Musicians who jammed together are grouped together in the figure, i.e. Sean, Andy and Zoe all jammed together in group 1.

Figure 5.8.: Touch scatterplots for the Viscotheque v2 jam session. The different colours represent which 'mode' the musician was jamming in (i.e. which parameters were mapped to the x and y axes). A description of the parameters is given in table 5.1.



(a) Parameter value distribution over all musicians and groups.



(b) Parameter value distribution broken down by group.

Figure 5.9.: Violin plots showing the distribution of the different audio parameters. The parameters mapped to the x touch position are in the top row, those mapped to y touch position are in the bottom row. For more detail about the meaning of the different parameters, see table 5.1.

5.5. Musician interviews

As in the v1 post-jam group interview the VCR interview was an open-ended one, and the musicians were encouraged to talk about any aspect of the experience they felt was meaningful. The key difference in the v2 group interviews was the use of the video recording as stimulus material for the discussion. The participants could ask for the video to be paused, and were able to comment in real time on their jamming activity. The participants watched the video on the same screen which showed the visuals during the jam session (see fig. 5.1). The video was shot in such a way as to allow the musicians to see both themselves (from behind) and also the visuals on the screen).

The interviews were transcribed in full, and nonfluencies (um, ah, like, etc.) were included following Potter and Hepburn (2005). These transcripts came to a total of 9340 words over the three group interviews. In this section, some excerpts from these interviews are provided. In these transcripts, **INT** denotes the interviewer (myself).

Regarding the issues raised by the musicians who participated in the Viscotheque v1 jam session, there were some similarities and some differences in the feedback from the musicians in the v2 jam session.

Novelty

The same excitement at using a new music making device was expressed by the musicians (and none of the musicians had used the v1 interface). This novelty factor manifested itself in a desire to explore, to test the limits of the mapping and their effect on the output sound.

Groove

There were moments of 'grooving' or 'being in the groove' described by the musicians. These moments will be discussed in more detail in section 5.5.1.

Legibility

The musicians described moments were they were not aware which sound they were controlling in the mix, although the practice session and visual feedback was helpful in this regard. As in v1, the timbral control parameters were cited as more legible than the rhythmic parameters.

Expressiveness

In contrast to the v1 instrument, a lack of expressivity (a feeling of limited sonic possibilities) was not mentioned by the musicians as a limitation of the instrument. While this may be simply because the issue did not come up in the jam sessions, the fact that it was not raised does indicate that the greater expressivity of the v2 interface was appreciated.

As well as these comments addressing the themes from the version 1 jam session, there were two overall themes which dominated the version 2 interviews: the relationship between groove and stability, and the different engagement networks which occurred in the jam session.

5.5.1. Groove

Overall, the musicians described the Viscotheque jam session positively, using words such as 'engaging', 'immersive' and 'fun'. The most satisfying and engaging moments of the performance were described independently by several group members (across different groups) using the language of *being in the groove*.

Watching the video of the jam sessions, there were definite moments of headbanging and toe-tapping. One such moment happened approximately 12 minutes into group 3's jam session, during an uptempo (i.e. small values of the *loop period* parameter) segment in which all three musicians' sounds making fast and loud. Upon seeing this section played back on the tape the musicians the musicians had this to say:

Ryan I also felt that, I've been, when you go through the whole digital world, web-based communities sort of thing, and one of the things I've realised is that a lot of it's coming back to the physical again, so here, for example [points at screen], even though the sound's really...the sound starts to become this communal groove, the points where we were most connected were the points when we were actually physically moving

Ted Mmm!

Ryan [miming headbanging] so **Ted** would be, like, headbanging, and I'd be headbanging, and...[points at screen] See! **Ted** is headbanging there, and it gives you that experience, it has to do with that as well.

Ted Yeah.

Ryan There is that human component, that's really strong.

Ted Because it does touch you

Ryan yeah

Ted when it's right.

This relationship between physical movement and a feeling of communal groove is consistent with the ideas of groove and entrainment discussed in section 3.2. More than just an individual physical feeling, though, **Ryan** points out these moments as engendering a feeling of connection and community.

There were other moments of physical agitation visible across all groups, but this moment exhibited the clearest and strongest co-ordination. When asked to explain why this part worked so well, the musicians explained

Ryan I sort of got really stuck into the rhythmic part, it was like, just trying to establish something—not a foundation, but a consistent stream, so that...if you were playing the drums, if you changed the tempo, it would just mess up everybody else's. So it was sort of like, that, adopting the role of the actual instrument, so, like, the guitar goes up and down and does these [mimes guitar playing]...the piano as well...but you know the beat has to be there, so it's like [claps rhythmically] 1...2...3...4..., 1...2...3...4..., 1...2...3...4...And I really felt embodied by that particular sample—having that role.

Ted So the role almost defines what you can actually do

Ryan Yeah

Ted As in, like, I mean coming from a musical background, the rhythm section does keep things going, and they feed off everyone else and provide the support, and they're quite mindful of that, but when you have a soloist, they're going all over the shop, like, they know what everybody else is doing, but they sort of do their own thing.

The assumption of musical roles from other group music contexts is unsurprising given the fact that the participants were musicians. In this excerpt, **Ryan** identifies with the rhythmic role of the drummer, as distinct from the more melodic roles of the other musicians. The assumption of certain roles by the musicians is also interesting because these roles were not imposed by the Viscotheque instrument.

Ryan's claim to have "really felt embodied by that particular sample—having that role" is an intimate way of describing a connection with the sound being produced. The power of rhythm in sound to "touch you", as described by **Ted** in the previous transcript excerpt, is similarly evocative. These comments regarding a close relationship between a musician and their sound is interesting in light of the indication from the data (in section 5.4) that the direct timbral control parameters (rather than the rhythmic/loop control parameters) were most utilised parameters by the musicians. However, a consistent rhythmic pulse was possible only when the loop period and offset parameters remained constant. It is not necessarily the case, then, that the musicians avoided these rhythmic parameters because they had no effect on the groove of the jam session, but perhaps that they had *too much* effect—they had the power to create and destroy the feeling of groove. Again, from group 3:

Ryan I also found that, for example, the 'rate/offset' variable tends to really throw off the coordination. So you'll be getting a nice groove going, and then [makes guttural sound], it stuffs it up.

Ted 'Cause it really is about the groove.

INT So tell me about that communal...tell me about that sensation of the groove that you were mentioning...

Ted It's just a beat that resonates inside you, that feels right. And it has to be coordinated, even if it doesn't all happen at the same time then [claps]...I'm not sure. And even now [refers to screen], we sort of feel a groove happening

Ryan Yeah.

Ted [makes rhythmic hand gestures in time with the music]

Ryan We're a band now!

ALL [laughter]

Ryan Bust out those iPods, we're going on a tour.

The comment about 'being a band', demonstrating real enthusiasm about the jam session and jamming group, is striking given that the musicians did not know each other beforehand. It is encouraging to observe this quality of interaction even after a short period of jamming with the Viscotheque DMI. These moments are similar to the sublime moments of flow which characterise and drive jamming in other contexts as described in chapter 3. While these specific and limited interview transcripts do not show definitively how and why these moments happened in Viscotheque v2, they are at least encouraging in demonstrating that they are *possible*.

The other issue which consistently accompanied any talk about the parts of the jam session that 'worked' was a discussion of the fragility of these moments. On several occasions the musicians described moments where they were *trying* to fit with one of the others, but were unable to because the other musician would change their rhythm.

Will I think it might have been around this time that I was trying to sync up the beats between—is yours top left, whose is that?

Joe That's me.

Will Yeah, between yours and mine.

Joe It's funny, because sometimes you move yours to sync with somebody else, but they've got a completely different thing in mind [laughs], moving it in a different way, or something...

⋮

Will It's kind of coming together a little bit

Dan That bit before was really different to everything else, which I think was good, it was a new idea

Joe Yeah there's some points where it really feels like we're all climaxing to one thing and then it's virtually impossible for us to all change to something else at the same time and for it to sound good.

INT So do you feel that that was one of those climax moments?

Joe Ah, yeah.

Again, lack of individual control over the whole is one of the hallmarks of the collaborative emergence which Sawyer sees as a key characteristic of jamming (see section 3.1). The overall sound is so fragile because each musician is subject to the whims of all the others, and it is difficult to know in unstructured jamming where each musician intends the whole group to *go* musically.

This is another example of the metaphor of conversation being apt for talking about jamming. Rather than a monologue, in which the speaker expresses themselves in a long chunk without any interaction or feedback, conversation involves short contributions with immediate feedback, including indications of agreement or disagreement and requests for clarification. In jamming, as in conversation, if participants don't listen to one another then the conversation goes off the rails. Indeed, it is always at risk of going off track, and only by the intentional interaction of all participants can it remain coherent.

The musicians in the Viscotheque v2 jam sessions repeatedly described the frustration of having other musicians change their sounds (particularly their rhythms) while they attempted to fit their own sounds around that rhythm. This is not a problem with the interface specifically, but it does raise the question of whether digital instruments can provide musicians with control dimensions or feedback which can help them avoid these frustrating moments.

Sean, from group 1, found the overall experience particularly frustrating. In particular, the lack of listening in the group led him to feel that there wasn't a level of 'respectfulness' in the group that is required for good jamming.

Sean I think there's a massive difference between enjoying what you're currently making, and maybe you're just excited that you're making a sound, but if you went to bed and came back the next day and actually listened to this on a radio you'd be like—what is this? And you wouldn't actually enjoy it so much.

This is perhaps the novelty factor at work in a destructive way—where musicians are enamoured with the sound making process simply because it is new, rather than sensitively trying to apply their music making skills to the new instrument. This is indeed a challenge in DMI design, but hopefully one which is mitigated as time goes on and the devices become more familiar. This was one of the reasons why the next series of Viscotheque (v3) jam sessions were held over several weeks, with multiple jam sessions for each group.

To summarise the idea of the groove as discussed by the jam participants, it was rhythmically (rather than timbrally) driven and it created a sense of community and satisfaction when it did happen. Group 3 in particular talked a lot about this feeling, but it did come up in the other groups as well. The negative side of the groove was its fragility—while setting up a groove required the cooperation of all the musicians, tearing it down required only one musician to start playing something which did not fit in. The was a need for communication and listening amongst the musicians, and all groups felt that these were areas in which they could improve.

5.5.2. Engagement networks

The term 'engagement' can be broadly used in two different, but related, senses (Peters et al., 2009). Engagement can refer to

1. the initiation of an activity
2. the state of being occupied in, or involved with, a given stimulus or activity.

In the context of Viscotheque, the two senses of the definition of engagement are related. The state of *being involved* occurs through the *initiation of activity* using the system. The sensation of 'being in the groove' discussed in the previous section falls into the second of the categories listed above. This state does not usually have a clearly-defined temporal begin and end.

However, the VCR technique also allows musicians to comment on *specific moments* in time. In the VCR interviews, the musicians indicated (while watching the tape) moments where they were attempting to deliberately respond to and fit in with the sounds the group was making. These specific moments of conscious engagement, indicated by comments such as 'I was trying to do this', or 'I was responding to that sound that [musician] was making'

Individual engagement	26
Unilateral engagement	18
Bilateral engagement	10
Total	54

Table 5.5.: Occurrences of each type of engagement code in the VCR interview.

were coded into three types of engagement: individual engagement, unilateral engagement and bilateral engagement. A breakdown of the occurrences of these codes is shown in table 5.5. Because the each group's interviews was a different length and covered different themes, comparisons between the groups are not meaningful and have not been provided. The purpose of providing these counts is to give an indication of the relative occurrences of the different codes in the interview analysis.

Individual Engagement

The experience of individual engagement (26 mentions) was characterised by a focus on the musician's own musical effects. This experience was described using words such as 'immersion'. This experience was reported at various times by almost all of the musicians, with **Sean** from group 1 the only exception, claiming

Sean So in the whole 15 minutes I don't think once we sort of achieved any band play together that was synchronised.

Often, the described moments of engagement involved the pursuit (and achievement) of high-level aesthetic goals:

Ryan For me it was all about strategy, see—here's the bit that I really enjoyed [points at screen]. So I was sort of finding the sound, and then I'd jump to number [mode] 5, and then I'd go back, high note...play around with the high note...and then I'd go back...and then a low note, and then a high note, and then (makes circular gesture with hands)...Strategy, you know? It was all about strategy.

However, this type of *individual* engagement was described by **Sean** as a cause of frustration—when a given musician's individual engagement with what they are doing renders them oblivious to the actions of the other participant and the collective sound:

Sean So if there were three people in the band, and there's one person just enjoying their socks off because they're just going off, but that is completely going in the wrong direction compared to the other people, well then that [enjoyment] may not be the case for the other people, so then they [the engaged participant] might enjoy their personal output, and not sort of be aware as to the outcome of the collective sound together.

Individual engagement is unique amongst the three types of engagement reported here in that it only involves one participant—the musician is simultaneously feeling engaged and responsible for the object (their sound and video) of their engagement. The fact that the musician is part of a group activity is irrelevant, their current engagement state is completely introspective. This Individual engagement is *not* group engagement, even if the individual is participating in a larger group activity.

Unilateral Engagement

At other points during the jam session musicians reported a conscious period of following or responding to another musician. This type of engagement was mentioned 18 times during the interviews.

> **Will** I think at this point I was trying to figure out how I could make the volume of my sample kind of play around with the volume of **Dan**'s. Like to try and get one to dominate at one time, then to dial mine back, the bring it up when his is going back down.

This type of engagement is a *unilateral engagement*—in these cases the musician whose sound was being responded to was unaware that they were the object of this attention. This is not necessarily a problem, musician A can follow the actions of musician B even if musician B is oblivious to the fact that they are the object of A's attention. It may even be the case that musician B is individually engaged in their own activity, unaware of *any* of the other musicians—musician A can still allow B's actions to influence his own. For this reason, unilateral engagement is also fragile, as discussed in the previous section on groove (section 4.5.2). In the binary relationship between the *subject* and *object* of engagement, the direction of engagement *matters*.

In contrast to the experience of individual engagement, during unilateral engagement a user was aware of and attending to the sound output of a fellow musician, rather than just their own. Unilateral engagement therefore precludes individual engagement. Here we see the potential confusion arising from a careless use of the word engagement—an experience can be engaging in completely different ways, depending on the nature and object of the engagement.

Several musicians also reported a deliberate cycling between the individual and unilateral modes of engagement:

> **Joe** I think I was trying to make a conscious effort to listen to other people, but there was times when I was trying to figure out something, and wanting to just do a thing myself.

> **Will** I think it kind of went through cycles for me, like I noticed there were particular points where I'd bring the resonance out, and just try and hear what my sample was doing, and there'd be other times where I'd go back to, say, the volume, and just bring it back a little bit and try and make sure that it was…kind of…a bit more subtle, I suppose, and going under the beat rather than trying to dominate it. So I suppose it was kind of backwards and forwards for me, but it was usually quite distinct in that I'd think 'OK, now I'm going to concentrate on what I'm doing', and then quite consciously switch back to 'now I'm going to try and make this work with the whole entity'

> **Joe** Mmm.

The type of engagement experienced by a given musician depended on which 'phase' of this cycle they were in. Each musician traversed a unique 'engagement trajectory' during the 15 minute performance period.

Bilateral Engagement

The third, and rarest, form of engagement mentioned by the musicians was *bilateral engagement* (10 mentions). This is when two musicians were consciously acting and reacting to one another in dialogue. Unlike unilateral engagement, in bilateral engagement *both musicians* are aware of the interplay. The distinction between unilateral and bilateral engagement is one not made in (Patel et al., 2009) or (Bryan-Kinns and Hamilton, 2009).

Unlike the other types of engagement, which were connected to specific parts of the performance by the musicians, bilateral engagement was more commonly referred to more generally, as some sort of vague *goal state*, or the sensation of being *in sync*:

Dan I really like this bit, or the bit that just went

INT What did you like about it, do you think?

Dan Well, it sounded like it worked, like it actually fit together...

Joe Yeah.

Dan We managed to get the samples working in a way that wasn't completely fighting with each other.

Joe Yeah, like with any sort of music I suppose, if someone wants to dominate, or lots of people are trying to dominate, it all sounds like rubbish, but because everyone's sitting back just tweaking and trying to be more subtle, I think it works.

The closest the musicians came to actually describing an exchange where they were wilfully playing *with* each other comes from group 2:

Joe I was trying to create a chord progression, but I wasn't [laughs] accurate enough.

Dan Yeah, I tried to do that too, I tried to go `I`$\rightarrow$`V`, and that's why I was trying to multi-touch.

Joe Mmm.

In this case, it seems that two of the musicians were trying to do the same thing coincidentally, rather than as a result of a conscious exchange of ideas. Even so, a sensitivity and balance between all the different musicians (which necessarily involves each musician being aware of the sound made by the others) was cited as a characteristic of the musical high points in each performance:

Zoe I think at this stage we were becoming more aware of each other. I think in the first little bit we were so concentrating...concentrating a lot on [hand gesture] our individual sound, and what we were doing, and how our manipulation was working, and at this stage perhaps we were becoming more aware of the other sounds, not that it consciously did that, but I think that's why it sort of, starts, maybe taking shape

⋮

Joe That sounds quite cool there, actually.

Dan Even there we seem to be able to, we're going with the same idea—which is interesting.

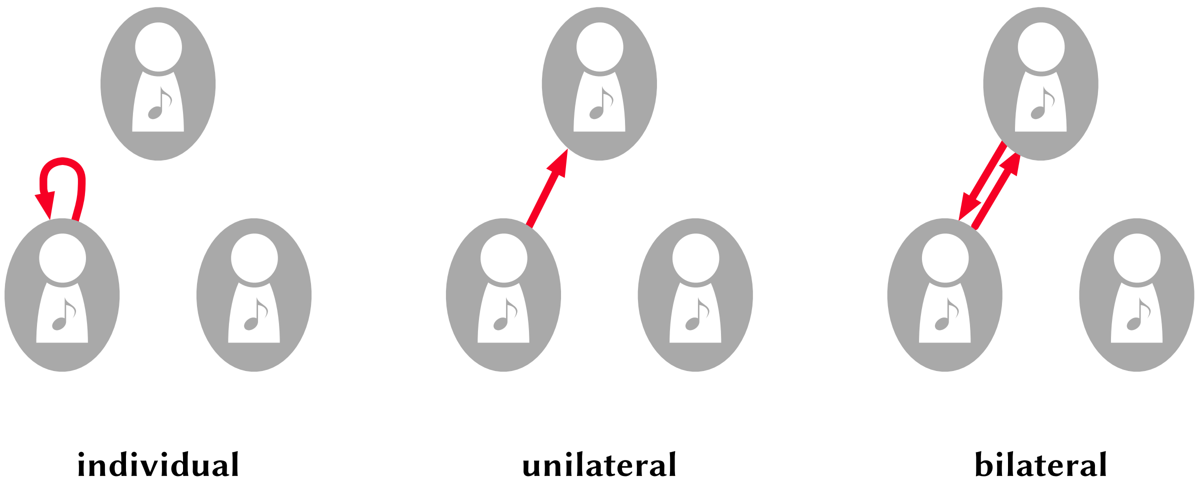


Figure 5.10.: Engagement networks representing the three different engagement relationships observed in the Viscotheque study.

The perceived 'quality' of the total musical output (which, while perhaps related to, is not the same as the depth of the musician's engagement) is here attributed to the harmonious synthesis of each musician's contribution to the larger whole. Of the three types of engagement, bilateral engagement seems the most likely to result in a truly productive awareness between the musicians, even if bilateral engagement was rare in the Viscotheque v2 jam sessions.

This tripartite structure is similar to other categorisations of engagement in multimedia, such as Bryan-Kinns (2004) *mutual engagement*, but is less specific—only concerned with the identity of the actors involved rather than the nature of their interaction.

Each of the three types of engagement was *felt* individually. At any given moment, each of the musicians may be engaged in a different fashion, or not engaged at all. Measuring the engagement of the group as a whole requires the construction of an *engagement network*, a directed graph representing all the different engagement relationships present at any given time (fig. 5.10). The distinction between unilateral and bilateral engagement is an important one. High degrees of unilateral engagement can occur without any bilateral engagement being achieved. When considering 'mutual' engagement, therefore, it is important to consider the *direction* of the engagement relationship.

If such a graph (or rather a time-series of graphs) could be constructed for a given jam session, it may be possible to comment meaningfully on the engagement dynamics over time in a given group. Representing engagement as a directed graph also presents the possibility of applying techniques from graph theory to the analysis of engagement networks. Average connectivity, cycles and other metrics can provide helpful insights into the engagement relationships during a particular performance. However, constructing these graphs is time-consuming (requiring in this case a VCR interview and subsequent interview coding process) and requires several judgement calls on the part of the researcher. It is not clear how this process could be automated in any simple way in jamming, due to the vast number of dimensions along which musicians may engage with each other, and the difficulty of discerning between real moments of conscious engagement and serendipitous accidents. It may be useful to consider more constrained interaction contexts, although such an interaction context begins to lose the open-endedness which is so crucial to jamming. For these reasons, automatic detection of engagement networks was not pursued in the next iteration of the Viscotheque DMI, although this may make for interesting future work.

5.6. Chapter summary

Building on the first stage of the Viscotheque design process, version 2 of the instrument provided a more expressive interface to the jamming musicians, both in terms of the number of parameters they could control and also in the move from discrete 'drum hits' to the ability to loop and morph longer segments of audio. This interface allowed the musicians more control over when and *how* the musicians sounds could be played, as distinct from the drum circle interface in v1 primarily giving them control of *when* their hits were triggered. This required the development of a custom iOS application and a redevelopment of the Viscotheque server from version 1. Also, a visual feedback element was added to the Viscotheque system to aid the musicians in orienting themselves.

Version 2 of the instrument was used by three different groups of musicians (nine musicians in total) for a free-form jam session. The feedback from the musicians indicated that there *were* satisfying moments of real groove. These moments were characterised by body movement such as toe-tapping as well as feelings of enjoyment and group identity, even though the musicians in each group did not know each other prior to the jam sessions. The rhythmic nature of the Viscotheque instrument, due to its use of a 'loop control' music making paradigm, was cited as a reason for the strength of the feeling of groove. The interaction between musicians showed some similarities to the hallmarks of jamming discussed in chapter 3, including the description of different objects of engagement in the jam sessions and the complex 'engagement networks' amongst the musicians.

In the log data, the musicians did exhibit some differences in their playing, both in the time spent manipulating the various parameters under their control and also in the values those parameters took. Overall, the direct or timbral parameters again proved more popular than the rhythmic control parameters, although this difference was less noticeable than in v1. A possible reason for this is the need to establish a consistent rhythmic pulse for grooving to occur.

There were also moments of frustration in the jam session caused by the musicians unfamiliarity with the device and at the participants not listening to each other. The issue of legibility (picking one's sound out of the mix) was raised as well, although it was not as large an issue in the post-jam discussions as in version 1 of the instrument. The shift towards a frustration with the other musicians rather than a frustration with the instrument itself is encouraging, and suggests that the increased dimensionality of the Viscotheque instrument as a control device provided the musicians sufficient scope for making music together.

6. Viscotheque v3

The third version of the Viscotheque DMI (version 3) was developed between March and November 2010. The key change to the Viscotheque interface in v3 was the use of a multi-touch sound mapping. The other main difference in v3 was the extension of the jam session/field trial of the interface to a month-long program of jam sessions. This gave the musicians more time to familiarise themselves with the interface and also provided the opportunity to see the way that their use of the instrument changed over time.

The v3 phase of the Viscotheque design process was the most in-depth of the three design stages presented in this thesis, both in the number of musicians and repeat jam sessions, and also in the analysis of the qualitative and quantitative data generated by these sessions. It is therefore covered in three chapters instead of one (as was the case with v1 and v2 of the system). The system description and experimental protocol are covered in this chapter, describing the architecture (section 6.1), mapping (section 6.2) and jam session protocol (section 6.3). The findings from the v3 jam sessions will be presented in the two subsequent chapters. In chapter 7 I will discuss the qualitative findings, including the interview themes and interaction patterns employed by the musicians. In chapter 8 I will present the quantitative analysis of the Viscotheque v3 log data, including representing the musicians' interaction with the Viscotheque instrument as a feature vector and identifying the musicians based on this data.

6.1. Architecture

The computing hardware (both server and iOS devices) used in Viscotheque v3 was the same as in versions 1 and 2 (see table 4.1). The rest of the physical environment was upgraded—a different jam room was used for the v3 jam sessions. These sessions were held in an acoustically treated studio space, which included a pair of full range Duntech DSM-15 loudspeakers. These speakers were capable of better sound reproduction than the KRK Rokits used for the first two versions of the system. In particular, the Duntechs provided greater bass response, both in terms of lower frequencies and reduced distortion. This was used to great effect by certain participants in the jam sessions, producing deep rumbling sounds capable of rattling glasses in the studio space. Using an acoustically deadened studio space also meant that there was less echo in the room, and that subtle reverberations were the result of the musicians manipulating their sounds through their instruments rather than artefacts of sound reproduction in the jamming space.

The seating configuration and location of the video screen were the same as in the previous jam sessions (see fig. 6.1).

The software environments in use for the v3 instrument were the same as in v2—Scheme (in Impromptu) for the server and Objective-C (on iOS) for the Viscotheque client applica-

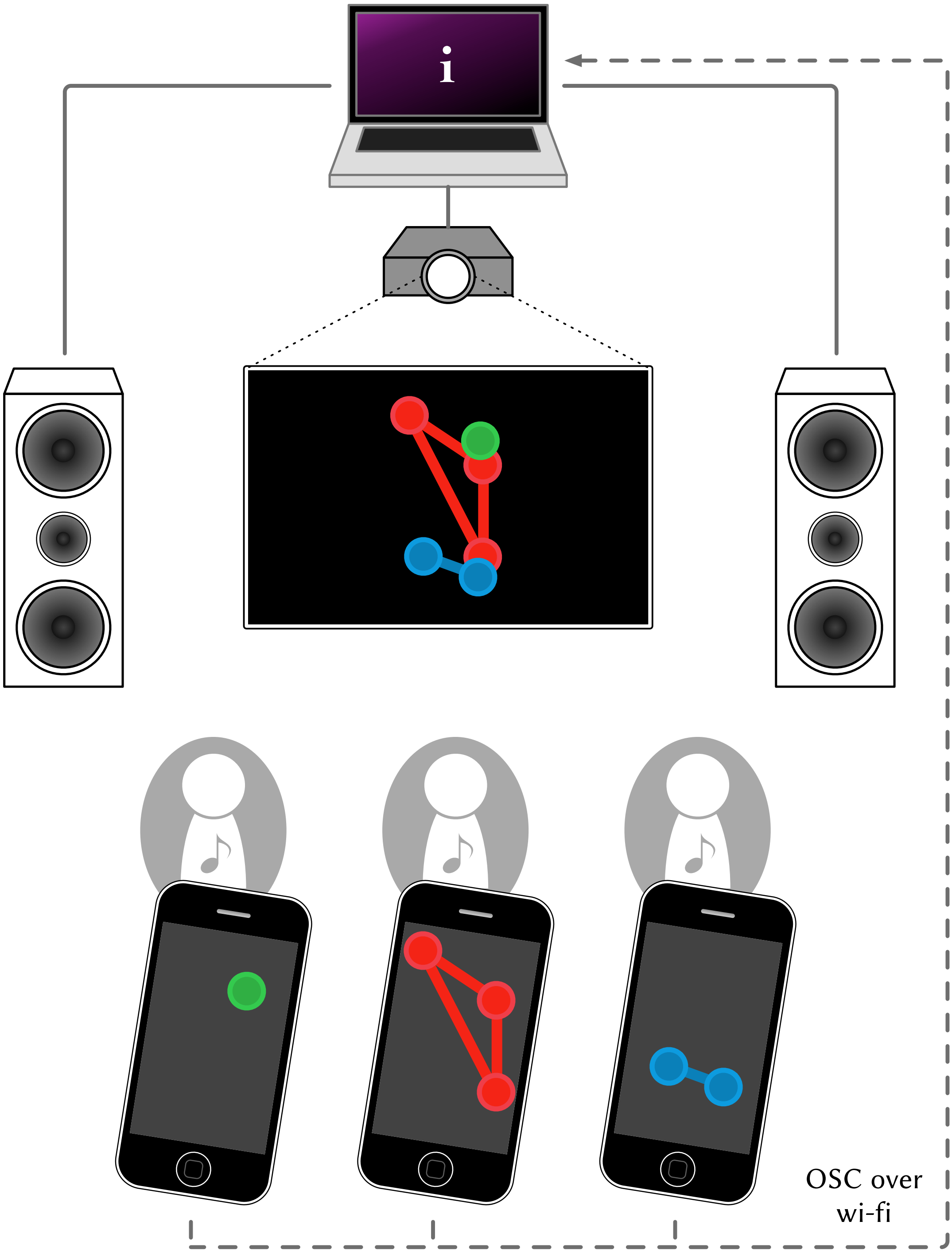
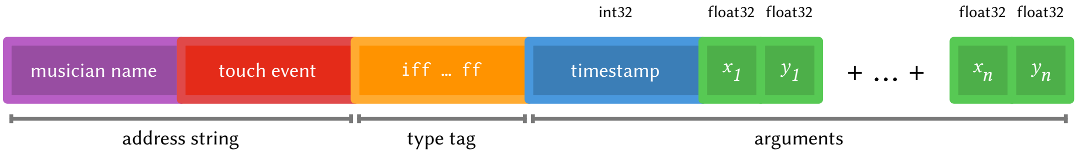
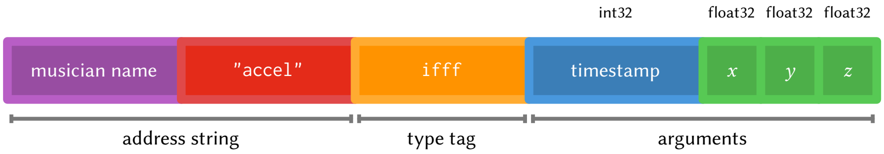


Figure 6.1.: Viscotheque v3 system architecture. The architecture was largely the same as Viscotheque v2. The primary changes were the addition of multi-touch input and the associated touch→sound/visuals mapping. There were still three co-located musicians seated in a semi-circular configuration facing the screen. A different jam room (compared to v1 and v2) was also used, with better speakers and sound damping.



(a) Viscotheque v3 OSC control message structure. The touch event portion of the address string will be one of {"up", "down", "moved"} as appropriate. See table 6.1 and fig. 6.3 for details about how these messages were mapped into sound.



(b) Accelerometer OSC packets. The accelerometer data did not control any aspect of the sound, it was collected to detect patterns in the movement and orientation of the device as the musicians jammed together.

Figure 6.2.: OSC messages in Viscotheque v3.

tion.[1] I rewrote both the server and the iOS application substantially for version 3. This was necessary to provide the updated audiovisual mapping at the core of the Viscotheque instrument.

The v3 OSC control message structure is shown in fig. 6.2(a). Unlike the control messages in v1 and v2, the data payload (and therefore the whole message) would vary in length depending on how many finger touches are on the screen. This was possible because each OSC message has its own 'type tag' specifying the number and type of the arguments, so that each message may contain a different payload (Wright, 2005, see).

Another addition to the Viscotheque system in v3 was the use of the iOS device's accelerometer to track the orientation and movement of the device as it was being played. The accelerometer was not used for musical control—the accelerometer readings did not influence the musician's sound. The data was collected to look for patterns in the device's physical orientation which might shed light on the way the musicians moved as they played the device. This accelerometer data showed some interesting similarities and differences between the musicians, and is discussed further in sections 8.1.3 and 8.6.

The accelerometer data was sent from the client to the server via OSC (see fig. 6.2(b)) at a rate of 50Hz. These accelerometer OSC messages were separate from the 'touch' messages (fig. 6.2(a)) which controlled the sound output. Both types of message were dumped to a log file by the server on receipt.

[1] The v3 server application was 1248 SLOC in scheme—a sixfold increase over the v2 server due to the more complex sound mapping. The Viscotheque v3 iOS application was 2522 SLOC in Objective-C (a 50% increase over v2).

Touch count	Manipulation	Effect
0		silence
1	touch down	begin playing sound
	x position	change low-pass filter cutoff
	y position	change low-pass filter resonance
2	y midpoint	adjust volume
	pinch/stretch gesture	time-stretch sound
3	y midpoint	pitch-shift sound
	pinch/stretch gesture	time-stretch sound
4	4-finger touch	toggle synth/sampler

Table 6.1.: The multi-touch mapping for the Viscotheque v3 interface (see fig. 6.3 for a graphical representation). For more than 1 touch, the parameter was determined by the midpoint value (i.e. for 2 touches the x midpoint is $\frac{x_2-x_1}{2}$).

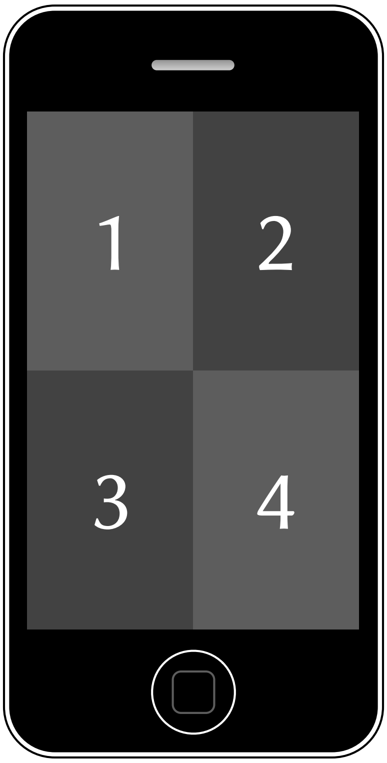
6.2. Mapping

Viscotheque v3 was a *multi-touch* sample manipulation and synthesis tool. The interface afforded pitch and time morphing of either the audio samples (as in v2) or a synthesiser. The synth and sampler were designed to respond in sonically equivalent ways to gestural input from the musicians—for example a 'pinch open' gesture would slow down the playback speed (using granular synthesis) when the sampler was active and slow down the oscillator frequency for the cross-modulation filter when the synth was active. This meant that there was a consistency between what effect a given action would have on the current sound regardless of whether the synth or the sampler was active. A four-finger touch on the screen would switch between the sampler and the synth.

As a sampler, the iPhone's screen was partitioned into four different zones, each of which triggered a different audio loop. Each loop was a short (4–8 seconds) audio clip of a single instrument (guitar, piano or drums—see fig. 6.3(a)) playing a simple pattern. The patterns were each one bar long, so that looping them produced a continuous stream of music with a constant pulse. The four different samples were not matched to each other—they had different tempos, accents and key signatures. This was by design, so that any coherence between the loops would be the result of effortful interaction between the jamming musicians.

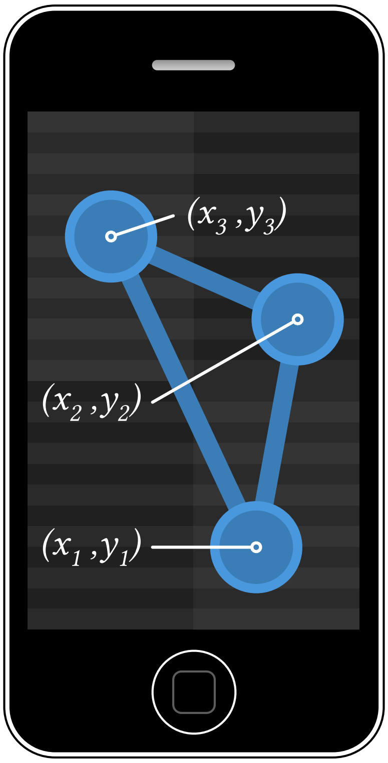
This mapping allowed for complex multi-touch gestures, potentially involving several fingers, affording the musician a large sonic range in which to play. The musician was in control of starting, manipulating and stopping their stream of musical material, potentially processing it to such a degree that it was unrecognisable.

When the sampler was active, touching the screen with one finger triggered the sample associated with that zone, and the sample continued to play on a loop while at least one finger remained touching the screen. Adding a second or third touch mutated the original loop



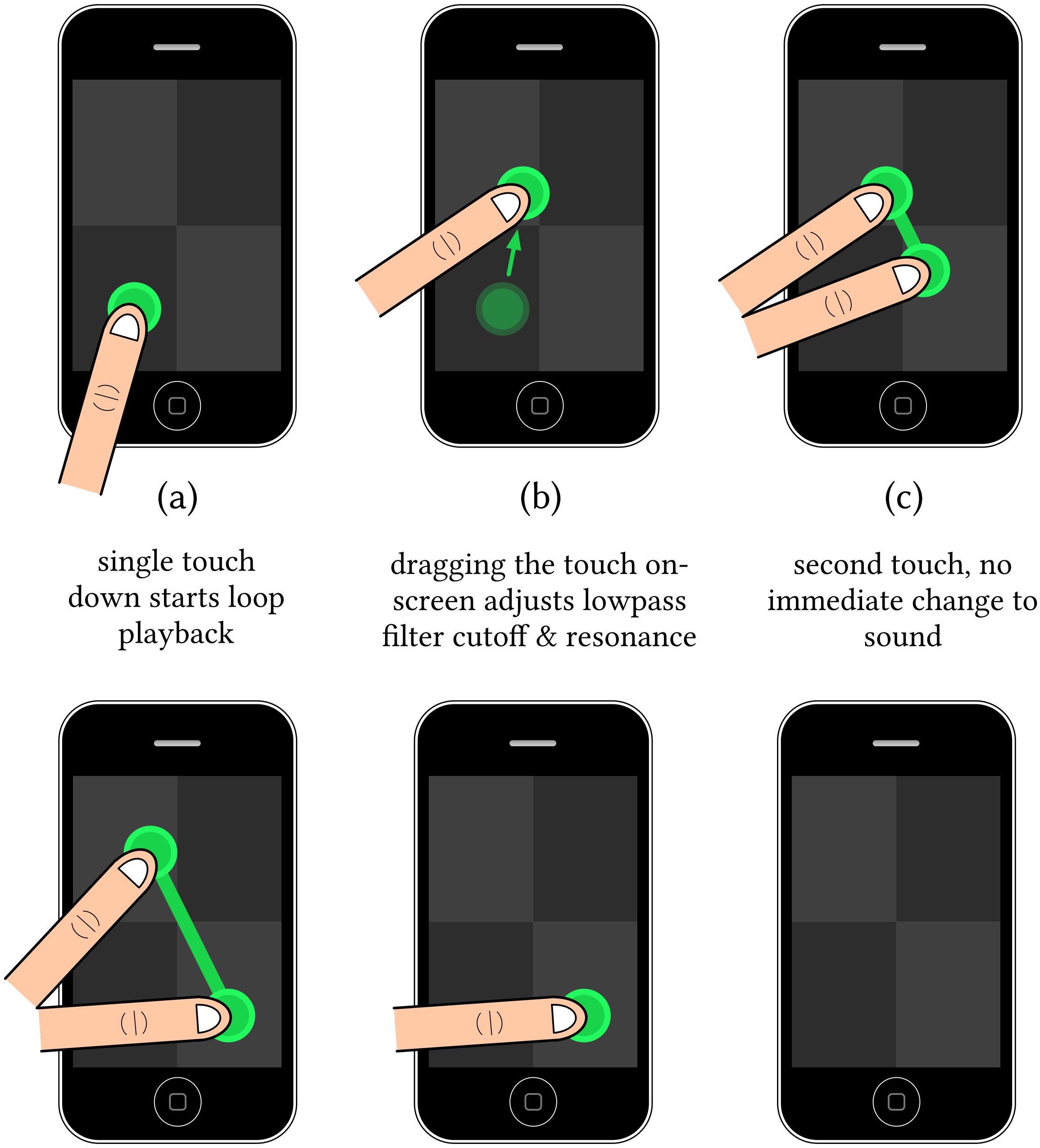
(a) The four sample zones. The *initial* touch determined which of the four samples was triggered. If that touch moved, or if additional finger touches occurred they did not trigger another sample, they manipulated the one already playing. In Viscotheque v3, the samples were:

1. drum loop (funk pattern)
2. Fender Rhodes ostinato
3. drum loop (16-beat rock pattern)
4. guitar riff



(b) Each finger touch was accompanied by a dot on the screen—the screenshot presented here shows three fingers incident on the screen. Each touch has an x and y position as shown, indexed by the order in which they occurred. Subtle zebra stripes on the background of the interface were provided to improve touch accuracy in the y dimension.

Figure 6.3.: Viscotheque v3 app interface. The mapping from touch→sound is given in table 6.1.



(a) single touch down starts loop playback

(b) dragging the touch on-screen adjusts lowpass filter cutoff & resonance

(c) second touch, no immediate change to sound

(d) loop continues to play, 'stretch' gesture applies slowdown effect

(e) second touch removed, loop continues to play (at slower speed)

(f) finger removed, loop playback stops immediately

Figure 6.4.: An 'touch storyboard': providing an example of a multi-touch sample playback and manipulation gesture on the Viscotheque v3 interface.

rather than triggering a second loop in parallel. When the last finger was removed from the screen, the sound stopped immediately, as shown in fig. 6.4. Similarly, when the synth was active, touching the screen triggered a sound with pitch determined by the y position of the touch. This 'touch-down→immediate sound' approach was a departure from the interfaces used in previous versions of the instrument. This removed the need for the loop period/offset parameters which governed the onset of the musician's loops—instead they could trigger a sound whenever they wanted by touching the screen.

Dragging a finger around on the screen or adding more fingers changed the processing applied to the sound. Up to four different fingers could be used at once (see fig. 6.3), and the effect the touches had on the sound depended on the number of fingers on the screen. When just one touch was dragged across the screen, a low-pass filter is applied. With two touches, the volume and the playback speed were modulated; with three touches, a pitch-shifting effect was applied. A full list of the parameters under the control of the musician and their operation is given in table 6.1. The mappings were designed to be intuitive, using conceptual metaphors wherever possible, such as 'up' and 'down' in relation to pitch and volume (Wilkie et al., 2010).

One key difference between the single-touch interface of Viscotheque v2 and the multi-touch interface of v3 was that the multi-touch interface was less 'stateful'. The v2 interface had five different modes which governed which audio parameters were mapped to the x and y positions of the touch. Knowing what effect a given touch would have on the sound required knowledge of what mode the instrument was in, and so button widgets were provided at the top of the touch-screen for displaying the current mode and facilitating mode-switching. In contrast, the multi-touch mapping used in v3 meant that seeing the finger touches (or being proprioceptively aware of the number of fingers on the screen) also revealed what mode the instrument was in, and therefore what sonic effect the movement of those fingers would have. This removed the need for auxiliary feedback on which mode the instrument was in, which was helpful not only for the musician themselves but also for the other musicians watching the feedback on the main screen.

The mapping was not completely stateless. The values of parameters such as the playback pitch and speed persisted even after the gesture which changed them was completed, so seeing the finger locations alone could not provide information about the values of *all* the parameters under the musician's control. Also, there the feedback did not provide information as to whether the synth or the sampler was active. Still, the move to a multi-touch interface did remove the need for explicit modes on separate interface 'pages' (as in v2), meaning that musicians no longer had 'remove themselves' from the sound to switch modes—each added or removed finger had a real (if subtle) effect on the sound.

The two main changes in the v3 mapping (immediate sound triggering and multiple touches) together had the effect of making the connection between the finger touches and the resulting sound more *direct.* Directness, in this sense (Flores et al., 2010, c.f.) means that any (and every) action had an *immediate* effect on the output sound, as opposed to a change that did not change the sound instantaneously (such as when modifying the loop start and loop period parameters in Viscotheque v2). Using the categories of section 4.2, this was a move away from a process control interface towards a more natural or direct interface.

This design direction was taken in response to the issues of legibility raised by the mu-

Participants	12 musicians, in 4 groups of 3
Jam sessions	4 sessions total, 1 session per group per week
Jam protocol	4×5 minute jams (20 minutes jamming in total) group interview (approx 30 minutes)
Data collected	touch logs (2.7M OSC packets, total log size 297MB) audio recording (direct stereo out from the server) video of jam session & interview interview transcripts (46271 words total)

Table 6.2.: Experimental overview for the Viscotheque v3 jam sessions. Unlike in the v1 and v2, in v3 the musicians came back for *multiple* jam sessions (four sessions over four weeks). This schedule is shown graphically in fig. 6.5.

Group 1	Group 2	Group 3	Group 4
Joe	Greg	Larry	Kate
Sarah	Leah	Tim	Judy
Alex	Alan	Chris	Roger

Table 6.3.: Version 3 jam session participants. These are not their real names, but the gender of each musician has been preserved in anonymisation.

sicians after the previous jam sessions. By creating a more direct connection between the musician's action and its sonic result, the musicians would hopefully be more able to identify both their own sounds in the mix and the sounds of the other musicians.

6.3. Jamming with Viscotheque v3

Aside from the more direct multi-touch sound mapping, the other main change in the Viscotheque v3 jam sessions was that the musicians came back for *multiple* jam sessions. There were 12 participants in these sessions, each one with at least one year of formal musical training as in the previous jam sessions. The anonymised names and groups for the v3 jam session participants are provided in table 6.3. Each group, having no initial experience with the Viscotheque DMI, attended four jam sessions over a four week period. A graphical representation of the jam session schedule is shown in fig. 6.5.

The groups were kept consistent over the four week period to allow the musicians to build a musical rapport. The musicians were not given any training or solo practice time in using the interface, although they could ask questions of each other at any time or of me during the post-jam interview. No instructions were given to the groups about what they were trying to achieve, although as musicians familiar with 'jamming' they brought with them their own expectations of what to do in an improvisational setting.

		Session 1 week 1				Session 2 week 2				Session 3 week 3				Session 4 week 4			
Group 1	Joe	J1	J2	J3	missing data	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Sarah	J1	J2	J3		J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Alex	J1	J2	J3		J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
Group 2	Greg	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Leah	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Alan	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
Group 3	Larry	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Tim	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Chris	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
Group 4	Kate	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Judy	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4
	Roger	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4	J1	J2	J3	J4

Figure 6.5.: The experimental schedule. Each musician was part of a group of three; these groups were kept consistent from week to week so that the musicians could build a rapport. Each week, the musicians would participate in four 5 minute jam sessions with a short (1 minute) break in between. After these four jams, the musicians were interviewed as a group.

The goal of the experiment was to see what patterns and cultures of use would emerge as the groups learned to jam together in Viscotheque. While the semi-controlled 'laboratory' setting opens the work up to criticisms of sterility and inauthenticity, there are significant advantages to being able to log every finger trace and capture every facial expression. More than this, though, the goal was to see how the musicians described the experience, and to see if the groups experienced the moments of deep satisfaction and euphoria associated with the best parts of improvisational music-making. How did the best bits happen, and what did they *feel* like?

The jam sessions were again recorded as detailed system logs—every OSC message (shown in fig. 6.2) was dumped to a log file. In addition, a video camera in the jam room was positioned so as to be able to see the faces of all the musicians, and the entire session (both jam session and post-jam interview) was recorded with the musician's consent. An anonymised still from one of the sessions is shown in fig. 6.6.

After the jam, the musicians took part in a semi-structured group interview. The VCR technique used in the v2 post-jam interviews was not used in v3 due to time constraints. The VCR technique can lead to long interviews, requiring at least enough time for the tape to be watched in full plus any time spent with the tape paused while the musicians discuss the experience. Given that the musicians were taking time out of their music studies, I wanted to keep each weekly session to an hour in total. By dropping the VCR interview in favour of a more traditional group interview, a better balance between *jamming* and *talking about*

Figure 6.6.: Three musicians during a Viscotheque v3 jam session. The left portion of the figure is the visual feedback which was being shown on the screen (which was out of shot to the left). I am present as an observer in the background.

jamming was possible while still keeping the session under one hour long (see table 6.2).

Interviews were conducted following Light's 'evocation' interview technique:

> The process involves questioning in a focusing, yet non-directive and unstructured, way. The interviewer chooses when and where to request more detail but leaves the direction, language and content to the interviewee to determine once the context of the interview has been set up and agreed (Light, 2006).

In the interviews, my questions were more focused on the experience of the jam session rather than the design of the instrument, but over the approximately eight hours of interviews the musicians also volunteered many comments about the design.

As in the v2 post jam interviews (see section 5.5), all interviews were fully transcribed and analysed for recurring themes. These 8 hours of interviews resulted in a 46271 word transcript. Over the four jam sessions (and therefore four post-jam interviews), the musicians were encouraged to discuss how their perceptions and experience with the instrument were developing as time went on.

Four weeks is a very short time over which to examine the evolving practices of jamming in with a new instrument, even for musicians trained in the skills and conventions of jamming. I hope to conduct longer studies in the future, and with more diverse users including non-musicians. However, this month of jam sessions represents the most substantial test of the Viscotheque DMI as a tool for jamming. The qualitative and quantitative findings from these jam sessions are presented in the next two chapters.

7. The affective power of sound

> I can *feel* it coming in the air tonight, oh Lord.
>
> *(*Phil Collins*, In the Air Tonight)*

One of the striking things about the v3 jam sessions was the atmosphere which characterised certain intense moments of the jam sessions. The room-filling sounds that the musicians produced at times led to some compelling moments—not compelling in an intellectual way, but rather in a way which was could be *felt*. The pitch-shifting and time-stretching capability provided by the v3 instrument led to the production of some wild electronic and synthetic sounds, far removed from their genesis as guitar and drum loops. These compelling sounds and atmospheres became more common as the weeks went by and the musicians became more familiar with the interface and with each other as a jamming group.

The video recordings of the sessions show encouraging signs of immersion and engagement between the musicians. At various points heads were bobbing, shared smiles were visible, eyes were closed—all good (although potentially misleading) indicators of the depth of musical connection and engagement between musicians.

In the earlier sessions, there are many obvious attempts to make music which was sonically coherent (at least to the western ear). As they familiarised themselves with the sonic possibilities of the interface, however, the sound did not always conform to what would conventionally be defined as music. The musicians at times created some dissonant and chaotic soundscapes. Some groups were more ready to embrace these sounds than others, but there was a general overall progression in this direction over the four weeks of jam sessions.

The interview transcripts were subjected to thematic analysis (Guest et al., 2011). In these interviews the musicians described a conscious effort to 'match' or 'fit' with the musical contributions of the others. However, the language used was not that of trying to fit with the other musicians, but of trying to fit with the *sound*. Rather than interacting with each other, the musicians described on many occasions the pull of the sound in guiding them and drawing them forward in their playing. The sound had a real power to *affect* the musicians, and to see the agency in these jam sessions as residing solely in the cognitive and bodily activity of the musicians seemed unsatisfying.

As a result, In this chapter I suggest that the concepts of *affect* and *assemblage* can provide interesting food for thought in the context of collaborative creativity in interactive systems. Some background on these ideas from the geography and cultural theory literature is necessary and is presented in the next section. In the rest of this chapter these ideas are used to explain the affective power of sound in the viscotheque v3 jam sessions.

This chapter showcases work already presented in the long paper B. Swift (2012a). "Becoming-Sound: Affect and Assemblage in Improvisational Digital Music-Making". In: *CHI '12: Pro-*

ceedings of the International Conference on Human Factors in Computing Systems.

7.1. Theory

The concepts of *affect* and *assemblage* proposed by thinkers such as Gilles Deleuze and Brian Massumi can help us to understand the interaction between users and artefacts in interactive systems, particularly in the context of computer-supported improvisation and creativity. Critical and cultural theory have a specific vocabulary, and some terms are used in subtly different ways to their use in traditional HCI discourse. In providing this background, I hope to address this potential for confusion.

Gilles Deleuze's thought (and that of his collaborator Félix Guattari) is renowned for its density and interconnectedness—both in form and in content. Deleuze initially rose to prominence for his unconventional readings of other philosophers, including Kant and Nietzsche, particularly his emphasis on difference and movement over identity and stasis (Smith and Protevi, 2008). His ideas have spawned a cottage industry of scholars interpreting and applying his work (Buchanan, 2004, p3). While he has been very influential in other fields, and despite the increasing popularity of cultural and critical theory in HCI discourse (see section 2.3) his work has had little impact in HCI. Satchell (2006, 2008) is the exception to this, although her reading of Deleuze does not focus on affect/intensity so much as his idea of the nomad. It is therefore necessary to provide a brief introduction to affect and assemblage theory.

7.1.1. Affect

Affect, or its synonymous term intensity (Massumi, 1995), describes the pre-personal, pre-reflective means by which all things (human and non-human, objects and ideas) affect one another, both positively and negatively. Affects are not the result of conscious processing and projection by human agents, they are the means by which bodies are empowered (or inhibited) to act, to *do*. According to DeLanda, bodies "possess an *indefinite* number of capacities to *affect* and be *affected* by other individuals" (DeLanda, 2002, p62, emphasis in original.)

The use of the term affect is problematic in HCI discourse. Picard and Cosier (1997) uses the same term in a slightly different sense, which is drawn from the psychology literature (Russ, 1993). In this tradition, the word 'affect' is used with a meaning very close to 'emotion'. In this sense, affective computing is about building computers which can sense and represent the affective-emotional state of their users (ibid.). This affect is biographical and personal. It is an emotion, felt and labeled, available to conscious introspection and reflection. This 'information processing' model of affect (and emotion) has been criticised for its individualism (Boehner et al., 2005).

This is *not* the meaning of the term affect as used in affect theory (see (Gregg and Seigworth, 2010)). Shouse describes the relationship between affect, feeling, and emotion thus: "Feelings are personal and biographical, emotions are social, and affects are pre-personal" (Shouse, 2005). A feeling is a sensation processed and labelled but still personal, while an emotion is the outward projection of a feeling. Affect is pure *potential*—abstract, unstructured and autonomous. It is this sense of the term I use in this thesis.

It should also be noted that there is a divergence of views on the nature of affect in social theory as well (in particular Tomkins, 1963). In this chapter I am presenting a view based on Massumi's (2002) conception of affect/intensity, which is influenced by his readings of Spinoza and Deleuze.

In *Ethics* (Spinoza, 1677), Spinoza is interested in the question of *what a body can do.* The use of the term 'body' is not limited to mean a human body—the concept is much broader:

> A body is not a fixed unit with a stable or static internal structure. On the contrary, a body is a dynamic relationship whose internal structure and external limits are subject to change. What we identify as a body is merely a temporarily stable relationship. (Hardt, 1993, p92)

Further reading on the history of the body in critical theory can be found in Blackman (2008).

Bodies are perpetually assaulted on all sides by many different affective/intensive forces. Some examples here may help to clarify things. Consider a chance encounter with an old friend—noticing them across a busy road. You cry out, wave your arms, perhaps jump between cars to cross and greet them. The body, responding to this recognition, is affected to do all these things, physically and emotionally.

Or consider a scenario where you enter into a room where two lovers have just been arguing. You may not catch any of the argument, they may revert immediately to an outward civility, but the atmosphere in the room is tense. In recognising this—in registering this affect—you may slink out of the room, or rise up and take a side in the argument, depending on your relationship to the couple (Brennan, 2004). These intensities are registered, enfolded and acted upon in different ways by different bodies, and affect is the name of the pre-reflective force which catalyses these actions.

> [Affective] atmospheres are the shared ground from which subjective states and their attendant feelings and emotions emerge (Anderson, 2009).

This is the autonomy of affect—it is 'registered' differently by different bodies, and affects which may arouse one may inhibit another.

In Spinoza's ethics, affects can be passive or active (Hardt, 1993, p100). Active affections are productive, they enable the body to act. Passive affects on the other hand are intensities that enfold a body, but only impact on the body's ability to feel or suffer. Furthermore, passive affects can be either joyful or sad. Joyful passive affects are a result of encounters with other bodies which are agreeable, consonant. Sad passive affects diminish the ability to act, they are a result of encounters with bodies whose internal relationships are not compatible with their own. These effects may be intertwined—affects may have both sad and joyful dimensions. The ethical project then, according to Spinoza, is to seek encounters with bodies which have 'an agreeable composition' to one's own body, with the ultimate goal of becoming active.

In affect theory there is no special place given to the human actor. The human body is subject to the affects which enfold it, resonating sympathetically as these affective forces pass fleetingly by as though just out of sight. The human is not an atomic, indivisible body, but is itself a composite body, with its own internal relationships and differences. This is at odds with the traditional primacy of the *human* user in HCI theory (as noted by Bardzell and Bardzell (2008)). This shift allows us greater freedom to understand complicated ensembles of digital artefacts, corporeal bodies, histories, and desires.

Music has an extraordinary power to shape our moods and actions. Malbon, describing his experiences in the UK club scene, writes

> the music and lighting effects combined so powerfully with the moving crowd on the dance floor...This kind of context—this sound and lightscape—must surely significantly change the ways that people interact. (Malbon, 1999, p. xii)

Musical sound can produce active affects, which may cause toe-tapping, singing or dancing. There may be sad passive affects at play as well, such as the depressive atmosphere created by the sombre horns of a requiem. As sounds enfold us we are affected; we are transformed. The nature of this transformation will be different for different bodies, depending on musical training, cultural background, current emotional state, and many other factors. The affective power of music to arouse the body is especially apparent in electrically amplified and digitally synthesised sonic environments such as the Discotheque (Lawrence, 2006), and also in the open-ended interactive systems of third wave HCI. This *intensity* of sound was evident in the v3 jam sessions and was the catalyst for seeking an explanation of this phenomenon in affect theory.

7.1.2. Assemblage

So what of this 'coming together' of bodies? In what ways can bodies come together to affect and to be affected? How does that shape our answer to the question of what a body can do? *Assemblage*[1] describes the organisation of bodies which opens up new possibilities for action (Deleuze and Guattari, 1987, Ch 4). The assemblage is not static entity, it is a process—a becoming, rather than a being.

> The assemblage is less about what it is then, and more about what it can do, what it can affect and bring about. (Dewsbury, 2011)

One of the key characteristics of assemblage thinking is a commitment to a flat (that is, non-hierarchical) ontology. A bicycle, a species of bird, a song, a mathematical theorem: these are all equally real, they are the product of the intensive forces which gave rise to them (DeLanda, 2002). Assemblages are scale-free—each assemblage may be a component in still larger assemblages (e.g. a person as a member of a family, which is a member of a society). A body of theory (such as HCI) is an assemblage—the result of different ideas, experiments and researchers affecting each other, sometimes strengthening sometimes discrediting, and always transforming.

Again, building on Spinoza's ethics of affected bodies, the assemblage is not directionless, it is striving towards new potentials. The progress of the assemblage towards these heightened capacities for action is not inevitable, and sad passive affects may inhibit this expression. When the assemblage resonates harmoniously, though—when its internal feedback loops

[1] *agencement* in Deleuze's native French

reinforce joyful and active affects—then the affective potentials are at their greatest, and the assemblage is empowered to become something new and different.

The assemblage is not a collection of identities but as a network of forces and intensities. And in the outworking of these intensities the assemblage is transformed, so that it can affect and be affected in new ways. This is perhaps the key difference between assemblage thinking and Latour's Actor-Network Theory (ANT) (Latour, 2005). Neither give any special place to the human agent in complex systems, but while ANT is concerned with what is required to *produce* a phenomenon, Deleuze and Guattari are more iterated in what possibilities for future *action* it opens up (Greenhough, 2011). The essence of the assemblage is in the opening up of new potentials for expression and action.

DeLanda (2006, p9) contrasts an assemblage picture with an organismic one—seeing a complex system as an organism (such as a human body). In the organismic metaphor, each component is works together in harmony to produce an organic *unity*. Examining any component of the system in isolation is problematic, because being this component in a larger whole is a core part of what it *is*. Also, as components are transplanted or repurposed in other systems (where their function is different) their identity necessarily changes as well. In an assemblage picture, by contrast, the nature of a component is fully defined by its intensive/affective history and the potentials for change that it opens up. If the component is 'plugged into' another assemblage, it may open up *different* potentials—affects are unstructured potentials, which affect different bodies differently if at all—but the component itself is not stripped of its identity in any way.

Another implication of this move away from the seamless unity of the organismic picture is the importance of heterogeneity in the assemblage. Differences between components are not to be glossed over, they are the animating force which drives the movement of the assemblage. The way a thermal gradient in a container of water produces a convection current, or genetic differences in a population give rise to diversity and adaptation through reproduction—these are examples of the *vitality* of difference. For Deleuze, difference is not simply the residual left behind when distinct identities are compared. Difference is prior to identity, and all identity and movement flows from the actualisation of these differentials.

This is a selective reading of the affect theory literature. This is obviously a necessity in a paper of this length. The Deleuzian corpus, with its rhizomatic and interconnected ideas, is sympathetic to this kind of treatment. Deleuze himself was known for his particular and sometimes selective readings of other thinkers (Smith and Protevi, 2008) and his appropriation of those ideas which he found productive in his own work. The ideas presented here can be useful in understanding the musicians in the Viscotheque v3 jam sessions as a jamming assemblage.

Each musician in an improvising ensemble is a body (indeed, is a composite body) with the ability to influence other bodies. They do this through their own musical contributions, as well as through their bodily movements and facial expressions. They are also vulnerable to the musical and bodily contributions of the other musicians. Each individual contribution must be understood in the context of the unfolding jam—prior contributions shape the way that present act of expression is understood.

Gilbert (2004) stresses the rhizomatic nature of improvisation. Unlike musical forms where the composer predefines the structures and relationships between the musicians and sounds ahead of time, in a jam the musicians are free to make their own connections, undirected by any central co-ordinating force. This is not to say that their interaction is chaotic or aimless. There are certainly creative forces at work in the improvisational assemblage, but they are decentralised. Any musician, through their music, can affect and be affected by all the others.

As already noted, affect is autonomous—it does not do the bidding of the bodies it enfolds. A musician can make a sound, they cannot control how that sound will contribute to the affects present and transform other bodies in its environment. These sounds are not affects themselves, but may give rise to powerful affective potentials, hinting at multiple potential musical and bodily responses. Rhythm, for instance, has the power to arouse and entrain movement in sympathy (Cummins, 2009).

The assemblage also provides an opportunity to critique the glorification of the human creative actor

> In place of the longstanding critical tradition that sees jazz through this lens of a metaphysics of human productivity, a range of actors, both human and non-human, come together in any given musical improvisation to construct a musical experiment. Instead of a protean, subject-based spontaneity, one discovers instead from this Deleuzian perspective, say, an instrument-club-musician-head-solo-influences-practice-time-mood assemblage. (Nesbitt, 2010, p159)

There *is* room for human creativity and intentionality in the jamming assemblage—the musicians are not leaves blown about on the winds of affect, robotically dancing to a tune they have no say in shaping. Rather, intentionality is simply an outworking of the nature of life itself: "life is always active and creative, affirming the power to become" (Colebrook, 2003, p66). One implication of the flat ontology is that while the musician *is* an assemblage of biological and experiential elements, they are not any less real or important than their constituent parts. So while there is no special place for a transcendent creative soul, the subjective feeling of intentionality experienced by the musicians is real—just as real as those interacting lower-level components.

The desire of the musical assemblage is to undergo transformation such that new potential musics are possible. Many factors may hinder or even dissolve the assemblage, such as a lack of instrumental skill or equipment problems. The fundamental drive, though, is to become active, to throw off any constraints which limit what sounds can be made. When we consider all dependencies of a musical improvisation event, all the factors which resonate in concert, it is no wonder that making sense of improvisational musical interaction is so hard and that articulating normative laws is so difficult.

7.2. Transcript analysis

The ideas of affect and assemblage discussed earlier in this chapter provided a productive influence in understanding the musician's reported experiences during the sessions. In this vein, our analysis of the interview data centred around two ideas: affective encounters between bodies, and the striving of the groups towards a 'becoming-sound'.

An empirical approach to studying affect must be sensitive to its prepersonal, elusive nature. It is a potential—it gives rise to action but is not action itself, therefore it can only be examined indirectly. To gain an understanding of gravity, one may drop a stone and observe it as it falls to the ground. In a similar vein, we observe the way that the bodies in the Viscotheque respond in response to the affects that enfold them. It is these 'affective traces' which are of interest in empirical investigations of affect. I am not trying to 'prove' or 'verify' these ideas, rather to see if they can help us understand this complex sociotechnical system.

This approach involves analysis of the notes taken during the jam session, the recorded audio-visual artefact, the video recordings of the sessions, and the group participant interviews. Glimpses of the affects at work may come from one or more of these sources—for example the facial and bodily postures of the musicians are shown in the session video while their subjective linguistic reflections come out in the interviews. My presence in the room during the sessions provided the opportunity to feel the affects first-hand.

The group interviews provided the musicians with a chance to reflect and discuss the experience directly after it occurs. In this section I hope to synthesise these multiple viewpoints to shed light on the jamming assemblage at work in Viscotheque. The data presented is a combination of observations from the video, 'readings' of the musical interaction and excerpts from the interview transcripts. As in section 5.5, in the transcripts the musician's names have been anonymised, and the sex of each musician has been preserved (female musicians still have typical female names, and the same for the male musicians). The breakdown of which musicians were in each group is shown in table 6.3. My contributions as interviewer are labelled **INT**.

Reflecting on one of their sessions, group 3 described a deep satisfaction and enjoyment reminiscent of that discussed in section 3.2.

Larry And then, and then you just, like, kindof recoup, and go back, and something—like there's points where there's something where it just all works, and for a second you just get that 'holy crap, let's just bottle this right now!'

Tim [laughing] Yeah

Chris Yeah

Larry Grab it, and just seize onto it, and figure out what exactly it is, because this is awesome.

Similarly, in group 2:

Greg For me, it's similar to other experiences I've had with other musicians, it's that moment of 'that's really cool', and yeah…it only really comes from playing music with other people, but it's like [clicks fingers] just a feeling where you go 'wow, that's clicking and that's awesome'. Yeah.

INT Do you think it can…

Greg It's something where you're working together, everyone's contributing to this really cool sound, yeah.

INT Yeah, sure.

Leah It was a lot more fun this week. Last week was more of a puzzle, trying to work it out, but this week it was a lot more free.

Again, in group 3:

Larry And yeah, all it takes is like someone to do something and I was like 'yeah, that's really cool, I'm gonna try this'

Chris Yeah

Larry and it's like 'yeah, that really worked!'

Tim Yeah

Larry Like when you [**Tim**] were doing dubstep before

ALL [laughter]

Larry it was like, oh crap—that's awesome!

Larry responded to hearing something cool by trying a new sound himself. We see here the improvisational nature of the assemblage—the response is an exploration rather than an assured composition, an experiment rather than a certainty. The success of this sonic experiment is shared by all, an interplay of forces which transform the moment, producing something 'awesome'.

The groups described moments of frustration as well. Sometimes this frustration was directed towards the interface, sometimes towards their inability to make the sound that they felt the musical context called for, sometimes and the group's unwillingness or inability to listen to each other and try and play together.

More than the previous incarnations of the Viscotheque instrument, the frustration felt by the musicians in the v3 jam sessions was more focused on *each other* rather than the instrument itself. This is encouraging from the perspective of the 'conversation of jamming', as discussed in section 3.1. The conversation of jamming with traditional instruments requires a balance of speaking and listening, of give and take (as discussed in section 3.1). That the musicians were able to feel that same frustration with the Viscotheque instrument hints that the interface was beginning to support real *jamming*, with all the associated benefits and challenges.

7.2.1. Affective atmospheres

The initial jam sessions were characterised by exploratory behaviour from all the musicians. Observing the participants on video, looks of concentration are evident on their faces as they play their instruments. This orientation period was characterised by frequent changes in visual attention; switching between looking at the instrument, the visual display, at each other, at the camera, etc. There was an air of flightiness or volatility to the jamming groups in these sessions.

The initial sessions were punctuated at regular intervals by moments of shared laughter and light-hearted banter between the musicians (e.g. **Sarah**: *"now I know how Björk's backing band feels!"*). Often surprised at the sounds produced by their own instruments, the musicians would regularly break from their own music making to share a smile or ask one another how a particular sound was made. The co-location of the musicians was a key factor here, as both verbal and bodily communication provided these opportunities for affirmation and interruption.

Musically, short bursts of sound with fast attack (the time between a sound's onset and its maximum volume) were used by the musicians in the initial sessions to identify their own sound in the mix. Rapid-fire staccato touches on the screen were also displayed on the main screen in the jam room. These sonic and visual factors reinforced this affect of flightiness and volatility.

So what did this affect of orientation, exploration and 'figuring out' empower the musicians to do and feel? **Joe**, after the first session, describes the mixture of interest, satisfaction and frustration felt during the jam.

Joe I think I would say that I was interested the whole time, but it sort of varied between whether it was frustrating or satisfying—but it was always interesting. Just trying to figure out in the early ones how it worked...in terms of working with my colleagues, but also figuring out what actually happens when you move this [points to the device] and how to access the different sounds.

The theme of 'figuring out how it worked' came up frequently in the interviews. While the valence of **Joe**'s feelings varies, the arousal—the pull to action—is a constant. Despite the frustration, the affective atmosphere in the early sessions is an active one, preparing the musicians for action.

As the sessions continued over the four weeks, some differences began to emerge, both within and between the groups. The volatility of the earlier sessions began to give way to a more stable, almost subdued atmosphere. Particularly notable was the way that the participants looked around the room, both at the screen and at each other. In the later sessions, these movements were less frequent and less obvious, although periods of head-bobbing occurred even when the sound had no strong rhythmic pulse.

Group 3 in particular was notable in the way the musicians in the later sessions would no longer look at each other, adopting slumped postures in their chairs, with eyes often closed. Along with this, there was a change in the general character of the sound being made, away from the choppiness of the earlier sessions towards smooth, sustained sounds, slowly waxing and waning together. The musicians, with their subdued physical appearance, were *less* obviously aroused, and to an observer could be mistaken for a group that was bored and uninterested. However, from the interview following group 3's third session (of four):

Chris ...and I did find that there were a couple of points where I was just 'wow, this sound I'm making is the shit', and I'm just having fun [mimes playing vigorously] and, well, there were a couple of times where I zoned out completely of what the other guys were doing

Tim Yeah

Chris 'Cause it's just—this sound is so wicked, I'm having so much fun with this—and you think 'well, hang on a second, I'm...this isn't just me here, I've gotta do something that they can work around, you know, so I can't be totally unpredictable.

This shift in atmosphere from one of skittish arousal to subdued calm was most obvious in group 3, but also to a lesser extent in the other groups. As I sat in the room in the later sessions, the change in affect even had an effect on me, as I shifted from constant note-taking to simply listening to the sound.

Notice that in the previous excerpt **Chris** is aware of the need to still work productively with the other musicians. In 'zoning out', **Chris** is worried that he may not be providing the other musicians something they can work around, impeding their ability to act (which is consistent with a theme in the v2 post-jam interviews—see section 5.5.2). His activity may, through network effects, result in sad passive affects elsewhere in the assemblage. Each Viscotheque jam session is a heterogenous assemblage of musicians, mood, experience, hardware, software and sound. Some of these interacting components work to strengthen and solidify, others to destabilise.

The musicians in group 1 did discuss these moments of inhibition and frustration:

Joe I think I would say that I was interested the whole time, but it sort of varied between whether it was frustrating or satisfying…but it was always interesting. Just trying to figure out in the early ones…figuring out how it worked.

INT And so what were the causes of your frustration, in general?

Joe Ah, I guess when things change all the time…

ALL Yeah [laughter]

Joe …in terms of working with my colleagues, but also figuring out what actually happens when you move this [points to the device] and how to access the different sounds. ⋮

Sarah I dunno, I tend to do the thing where I find something that's good, and I stick on it, or I think it's working, and I spend a lot of time trying to find that, I think, because it's, I dunno…you feel like you can't drop in on someone else's sound, so, you know, you've gotta find your own thing while you're doing this…it's kind of frustrating at times.

INT What aspect of it is frustrating?

Sarah Like, I was on the whirly thing (mimes circular motion)

Joe Making a fresh sound

Sarah and then M3 said 'try something else!', and I'm 'aargh! I don't know how!'

This discussion refers to an incident which occurs after a sustained period of synchronised and rhythmic musical activity involving all three musicians. **Joe**, perhaps feeling a need for the music to change, says *"**Sarah**, you change something now"*. **Sarah** describes feeling the pressure to not 'drop in on someone else's sound'. The current sonic context governs what sounds may follow at any given point—the sound powerfully shapes what the musicians can and cannot do. This feeling, combined with her inexperience with the instrument, renders her unable to respond to the instruction from **Joe**. A diminished capacity to act musically is combined with a feeling of frustration. This is a mixture of sad passive and active affects. A commonly reported cause of frustration is the hindering of sonic ambitions:

Chris I spent the next ten minutes trying to work out how you [**Tim**] made that sound

Larry Yeah!

Chris I was like, I've been playing this thing for four weeks, man, how did you make a sound like that?

One recurrent theme in the participant interviews was the musicians' compulsion to fit in with the sound, particularly as they became more familiar with the instrument.

Greg Yeah, and I think it gets—I think I know similar things in terms of—you get to that stage where you think 'oh, ok, I'm just gonna sit back a bit and try and slot in, rather than 'I'm just taken with whatever sound I'm doing'...try and work it. You get to a point where you try and work a bit more as a group.

The notion of 'slotting in', of acting coherently in the current sonic context, drives the actions of the musicians as they jam. In terms of governing what the assemblage can do, the sound is a powerful affective agent, shaping the complex interactions between the musicians, the instruments, and the environment. The sound creates an affective atmosphere which determines what *fits* and shapes the actions of the musicians. This is unsurprising—musicians have a deep affinity for sound, and come together to jam with specific expectations about the nature and composition of their interactions. The group's sound, the harmonious (or dissonant) blend of all their musical contributions, has perhaps the greatest effect on the affective atmosphere and unfolding behaviour of the jamming group.

This is in contrast to the romantic picture of fully self-contained creativity springing forth from an individual musical genius (Nesbitt, 2010). The sound here is far from an inert medium in which human actors carve their initials—it is an affective agent. Against the notion of a reified 'user', affect theory (and other 'posthuman' ideas, such as the cyborg (Haraway, 1991)) is more open to seeing the agentive elements in all bodies, ideas and processes.

7.2.2. Becoming-sound

Gilbert (2004) suggests that the improvising assemblage at its most active and most capable of expressing itself is characterised by a blurring of boundaries, with musician, instrument, and sound all moving together as one. The assemblage reaches towards a becoming-music, a harmonious and resonant productive flow of intensities moving together in musical production.

Tim Yeah, the best times are when we don't think about it,

Larry Yeah

Tim cause that's when it's most surprising

Larry And that's when you just put something down, and everyone would be, like—wait! And you could just feel this moment of starting into nothingness and playing with some fingers on this [mimes playing the device] and it would just all fit into place, I find.

Tim Yep

Larry There were definitely points—there was one, I can't remember if it was the third jam or not—but I was just doing something, and it ended, and I was, like [mimes looking at watch] we just started, we literally just started!

This absorption was also noticeable at the end of each jam session, when the instrument turned itself off and the sound stopped abruptly. This disruption of the atmosphere caused (at different times) laughter, swearing and audible exhalation.

In Viscotheque, this blurring of boundaries between musician, instrument and sound came through in the way the participants talked about their agency in making sound with their instrument. In the earlier sessions they talked about their sound as '*the* sound', however by the final session the language of '*my* sound' was becoming common. The instrument afforded the musician a presence and influence in the 'world of sound', a world of powerful affects and intensities. Becoming familiar or competent involves bridging the gap between the finger manipulations and the sound required by the current sonic context.

A key question, from an assemblage standpoint, is what are the intensive differences which drive the assemblage forward, opening up new potentials for musical expression? One of the interesting aspects of musical expression is the importance of repetition. The 'doing' of music, even in the simplest case, requires sustained activity from the assemblage of musician and instrument. Change, or *transformation*, on the other hand, involves embellishing, developing or destroying these patterns of activity. To examine 'what a body can do' in Viscotheque, then, is to observe the affective atmospheres which give rise to transformations and transitions in the jamming group.

One recurring feature of the jam sessions was the impact of sonic 'discontinuities', such as the introduction of an interesting timbre, a sudden loud noise, or the sudden removal of a sound. Some of these sounds were obviously unintentional and serendipitous, as evidenced by the expression of surprise from the musician upon making the sound. Others were a deliberate attempt to change things up, while still others were an attempt to blend in and fit with the current sonic context which was interpreted and transformed by another musician. When these moments occurred there was often a concerted effort from all musicians to fit with this sound and to produce a sound which was sonically coherent in the current context. Sometimes they were able to find such a sound, and sometimes they were not able to before the original compelling sound—the catalyst—disappeared, perhaps because of boredom or a lack of skill. When these moments of coherence *did* occur they often persisted for a short time (up to 60 seconds) as the musicians made subtle variations to their sounds in an attempt to develop the sound further. Then, after this time, the feeling of coherence (and the associated intensity) would disappear, either gradually dissolving or catastrophically breaking down. The moments novelty—of *difference*—were frequently the catalyst for the group as a whole changing the overall atmosphere of the jam.

Larry Yeah, I think what I enjoyed from it was the points when something would…you could just feel that little *click*, and it would just, you just kindof went 'bang!'—fell into this position, and it was like 'OK, this is it, we're here, we've got it'…

Tim yeah

Larry …and then it would just be, like, **Tim** would start doing this, just a little tap or something like that, and then it would work…

Chris yeah

Larry …and then **Chris** would just bring up something like that, and I would just, kindof, be messing with this thing, and it would all just accidentally fall into place.

Tim Yeah, I wasn't even *trying* to make it work, it would just work...

Larry ...and it was just experimenting, yeah. And then when it worked, or when we found something where we all linked, it was, like—'bang!', it was just, like, you know...a lion pouncing on a zebra, or something.

ALL [laughter]

Chris ...just flick the switch, it was like, 'bang!', it worked.

This is a striking and vivid description of how quickly the feeling in the group can change—a change registered by all the musicians. The sound is the affective agent, providing the unstructured potential which is then appropriated by the musicians in their response to it. The feeling in the room could change in an instant, as a new and different sound created a sonic atmosphere which almost *demanded* a response.

As the musicians became familiar with the instrument they showed a willingness to move beyond the conventions of western art music. Group 1 was particularly interesting in this regard. **Joe**, whose musical training is in classical voice, began the sessions playing fairly conventional melodies and rhythms, taking advantage of the responsiveness of the instrument to tap out repeating motifs. These motifs would often cause the other musicians to try and sync up, with varying degrees of success. As the sessions progressed, though, **Joe** began to explore the synthetic digital timbres the instrument was capable of producing. After the final session, reflecting on his surprise at the way his sound developed:

Joe I actually think, given what you have to use—which is four loops and a synthesiser—there's a lot more than I expected initially, there's a lot more potential than what I initially sortof assumed. And so I wouldn't say that I wasn't looking forward to it, but that I, um, yeah. And I think in the second session, when these guys started making some of the cool sounds you get when you slow things right down and that sort of thing the really different sounds to what you get to start with, then it was like 'oh, there's all these things that you can do'

From an assemblage perspective, this is the *real* opportunity and benefit provided by the Viscotheque instrument, and indeed DMIs in general—the ability to throw off constraints about what sounds can be made and bring new sounds within reach. Group 3 in particular were notable for the way they embraced the 'digitality' of the instrument and its timbral possibilities. Particularly in the later sessions a 'drone' atmosphere was created, as the groups eschewed rhythm (either in the form of triggering the sounds in any sort of rhythmic pattern or using the drum loops) in favour of loud, space-filling digital timbres.

The sound is not the *only* factor which contributes to the affective atmospheres in Viscotheque—the system's visuals, the musicians' bodily and verbal expressions, their mood and many other factors contribute as well. However, the affective power of loud sound and music is enormous. This has implications from a design standpoint, as music, video and dance are increasingly common features of third wave HCI. They create a high intensity environment which is unlike other human-computer interaction contexts like web surfing or word processing. The powerful affects these elements can contribute to an environment can have a significant impact on the behaviour of the bodies they enfold, and in these contexts affective atmospheres must be considered in the design process.

7.3. Chapter summary

This chapter mobilises a notion of *affect* that is something more than a subjective feeling. Sound and music possess a great deal of power to affect musicians and set up intense and heightened atmospheres in which creativity and action can flourish. These intense atmospheres were present at times in the Viscotheque v3 jam sessions, particularly as the musicians moved past simple loop playback of conventional instruments and towards the synthetic sounds afforded by digital signal processing. More than the previous two versions of the interface, the musicians described moments of genuine enjoyment and satisfaction in their jamming, as well as moments of frustration.

I hope that this chapter has provided an insight into how the concepts of affect and assemblage can be mobilised to understand the rich, open-ended, creative interactions in HCI that are so difficult to examine sensitively. As third wave HCI wrestles with issues surrounding creativity, play and self-expression, the affective dimension of these environments cannot be ignored. More work needs to be done to integrate these concepts into a mature design practice.

8. The differentiating power of data

It's in the way you hold me...

(Shania Twain,
You've Got a Way)

One characteristic of some third wave HCI has been a rejection of quantitative methods in favour of the thick descriptions of ethnography and the relativism of cultural and critical theory. As discussed in chapter 2, this push has some merit—creative, social encounters with technology *are* complex, and to examine them too narrowly is a pitfall worth avoiding. The previous chapter heeded this criticism in its use of a prepersonal, autonomous understanding of affect and the power that sound has to shape the musicians it enfolds even as they shape the sound through their instruments.

However, the digitisation of creative and cultural practices (such as music making) allows us to log, examine and reason about them more powerfully than ever before. Every keystroke and finger touch, every slight shake of the device—these actions are laid bare by the array of sensors in these devices. In the Viscotheque v3 jam sessions the touch, accelerometer and audio output data was all recorded in full.[1] To ignore this information outright seems wasteful.

The challenge, then, is to examine this data in a way that is sensitive to the third wave critique. In this chapter I shall explore how modern Machine Learning (ML) techniques can assist (rather than replace) the human observer in making sense of the improvisational group music making in Viscotheque.

The theses of this chapter are:

1. based on the log data there *are* differences in activity between the musicians, and it is possible to train a classifier to discriminate between the musicians based on these differences; and

2. the *nature* of these differences is as interesting as the classifier accuracy.

All data analysis has been performed using the GNU `R` Project for Statistical Computing software environment (R Development Core Team, 2012). All the figures were produced with the `ggplot2` package (Wickham, 2009). Other libraries are referenced where appropriate.

8.1. Data visualisation

Perhaps the most obvious advantage to having each musician's data in a machine readable form is the potential for data visualisation. In partitioning the data for visualisation, there

[1] Details about the data logged in Viscotheque v3 are given in fig. 6.2.

are two main ways of breaking the data down: by group, and by musician. Looking at the data by group can reveal differences between the groups as a whole, while looking at the data by musician can show differences between the individual musicians. In this section I will use both partitioning approaches.

8.1.1. Finger traces

A straightforward plot of the different touch patterns used by the musicians reveals some interesting differences between musicians and between the groups, as shown in fig. 8.2.

These 'finger traces' are an overall view of *every* finger touch on the screen. This representation flattens out the time dimension of the interaction and shows the parts of the screen favoured by each musician in their finger touches.

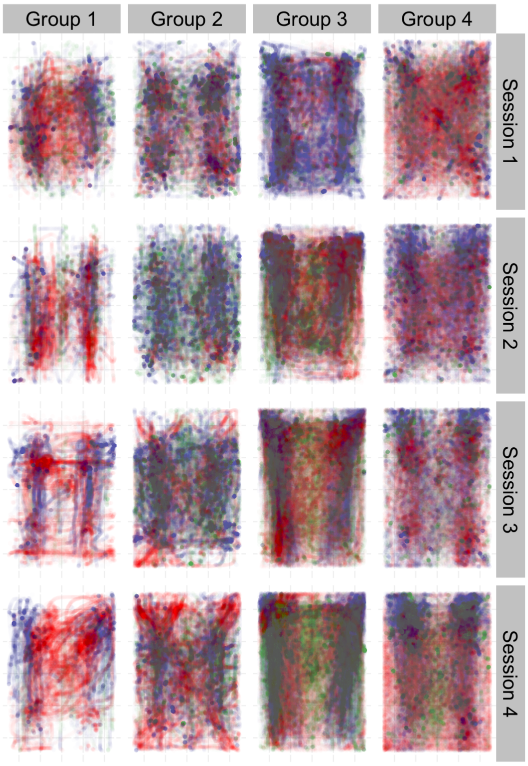
The first thing to notice in fig. 8.2 and fig. 8.1 is that there *are* noticeable differences between musicians and between groups. A few key observations:

- group 1 has the most sparse touchmap, group 3 the most dense
- the top left and right corners of the screen seem to be 'hotspots' (most obvious in group 3)
- there are a couple of recognisable motifs—such as the red (1-finger) 'X', and the blue (2-finger) vertical stripes on the left and right edge of the screen
- there seems to be more similarity between weeks than between groups/musicians—the differences between musicians are greater than the differences within musicians and groups over time

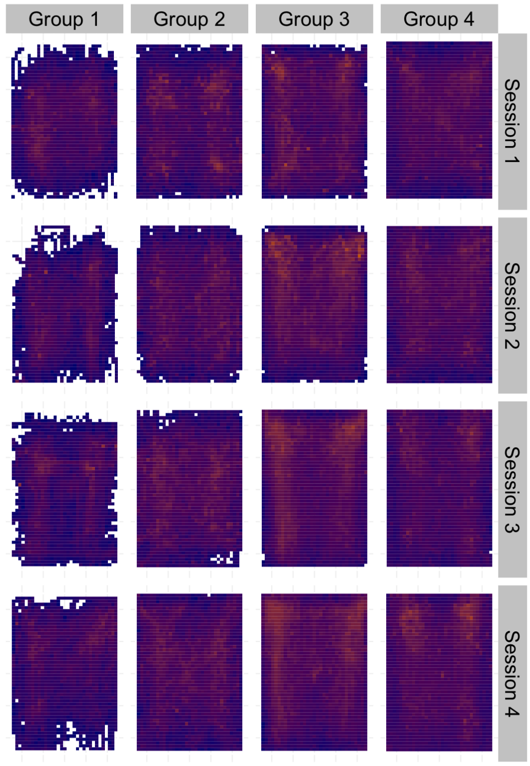
In fig. 8.3 the touchmaps are shown partitioned by touch count. The effect of touch count is twofold—it has a musical and a physiological basis. Each different number of touches corresponds to the musician playing in a different mode of the system, with a different musical effect. However, the shape and capabilities of the hand also have an effect. With two fingers, for instance, there are two prominent vertical stripes on the display. Watching the musicians on the tape, these patterns are caused primarily by the index and middle finger in a '*v*-shape' sliding up and down the screen. This gesture controlled the volume, as described in section 6.2.

The similarities between musicians in the same group should not be surprising. The visual feedback presented to the musicians as they jammed showed what the other members of their group were doing. This allowed for both straightforward copying and also subtler influence and inspiration between the musicians.

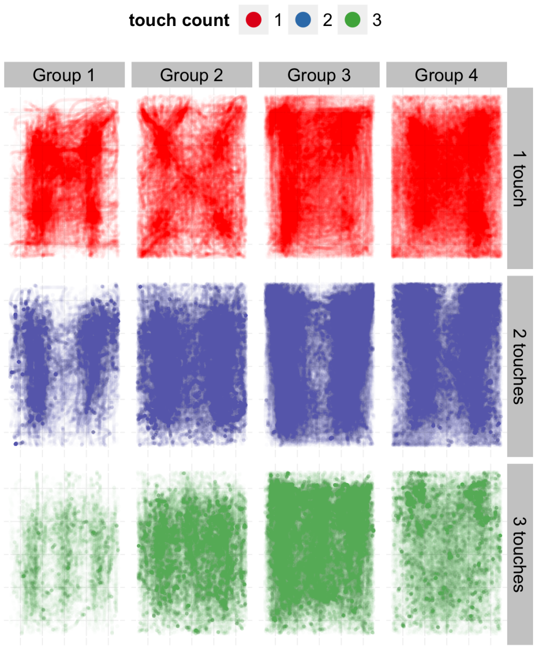
It is also interesting to look at the way these patterns evolved over the four sessions in more detail (fig. 8.3). Again, these plots seem to bear out the pattern that the inter-musician diversity is greater than the intra-musician (between sessions) diversity. In the figures this means that *in general* the columns show more similarity than the rows. This suggests that differentiating between the musicians based on their log data may be possible, and that what I shall attempt to do in section 8.2.



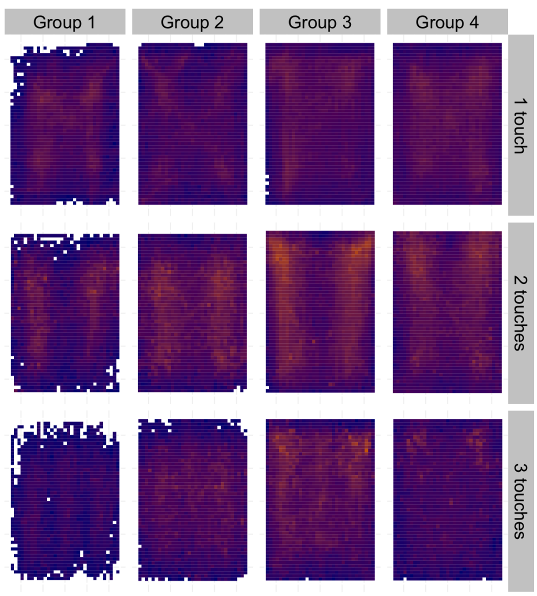
(a) Touch scatterplots by group.



(b) Touch heatmaps by group.

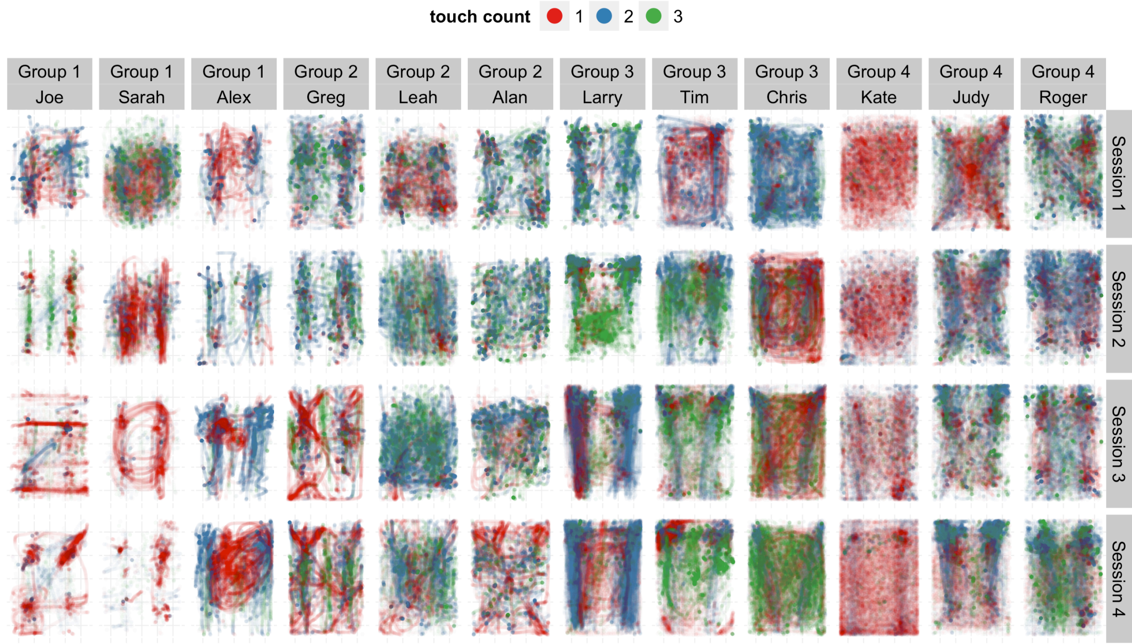


(c) Touch scatterplots by group.

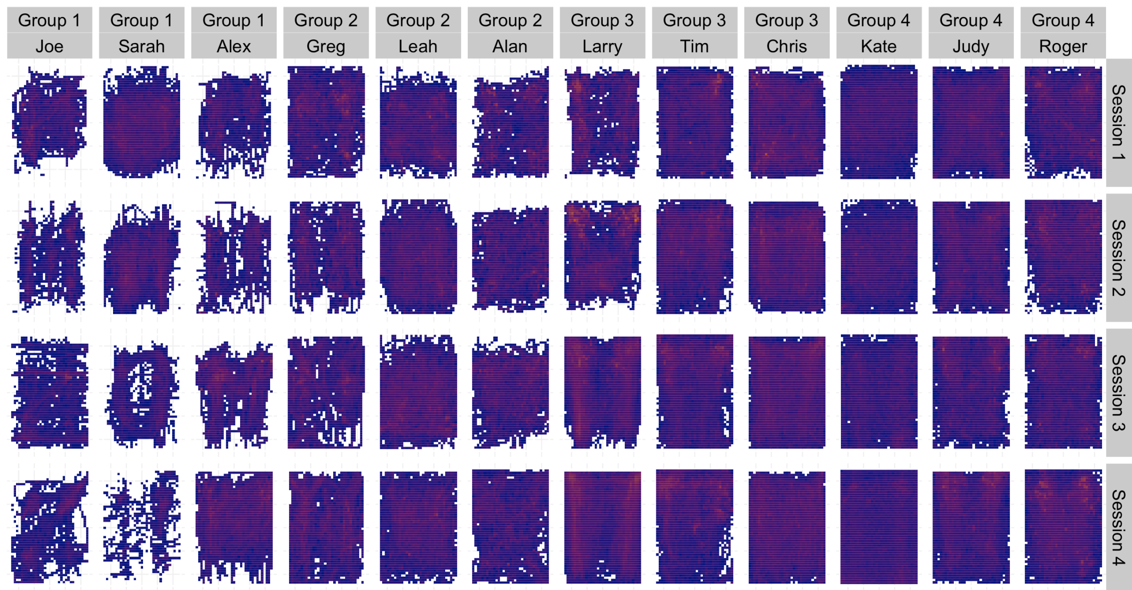


(d) Touch heatmaps by group.

Figure 8.1.: Total touch activity by group. The top two plots show the differences in touch activity over the four sessions, while the bottom two show the differences between the different modes (touch counts).



(a) Touch scatterplots by musician.



(b) Touch scatterplots by musician.

Figure 8.2.: Total touch activity by musician, showing each musician's development from the first session (top row) to the final session (bottom row).

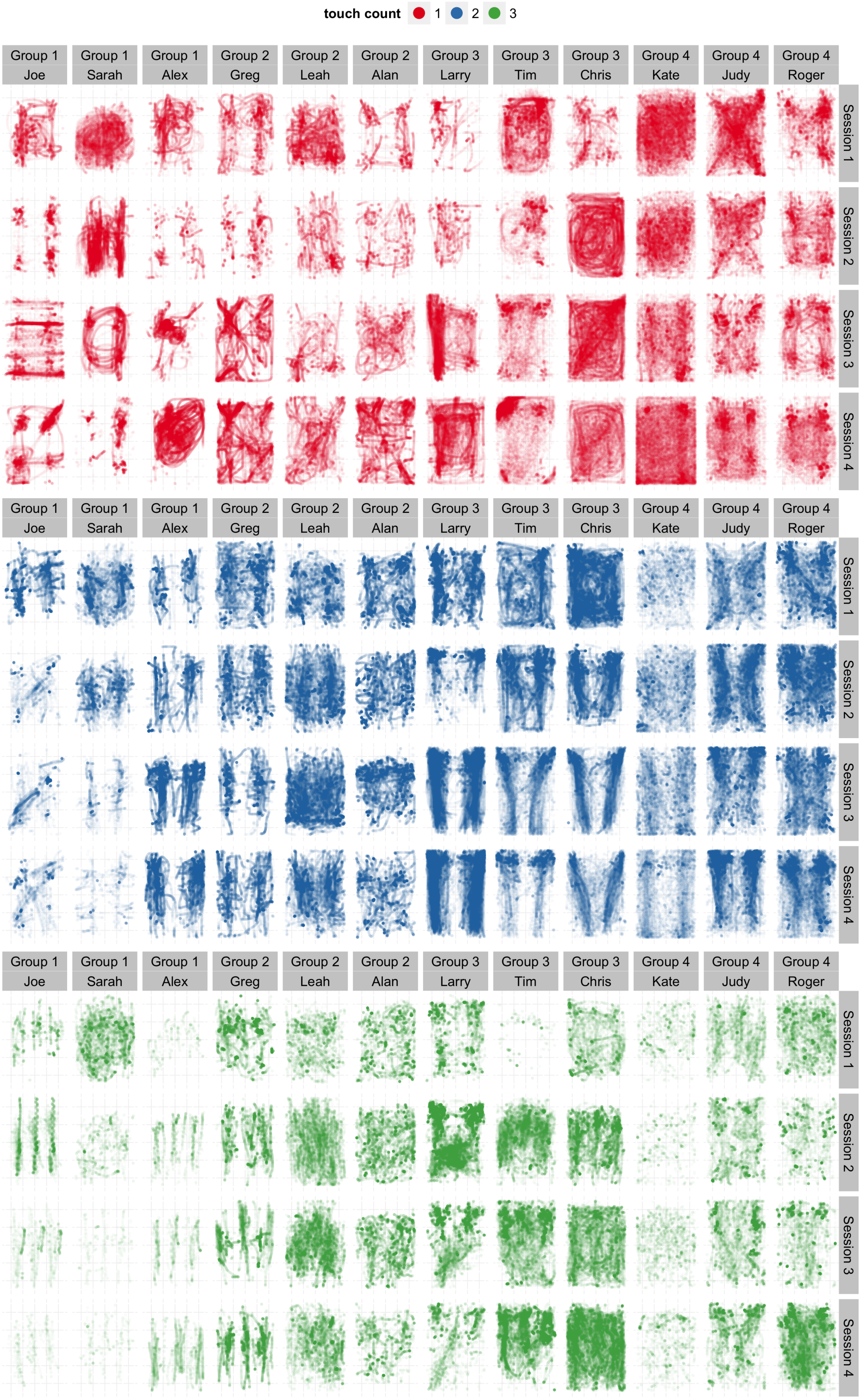


Figure 8.3.: Touch scatterplots over time, showing each musician's development from the first session (top row) to the final session (bottom row) for each of the different modes.

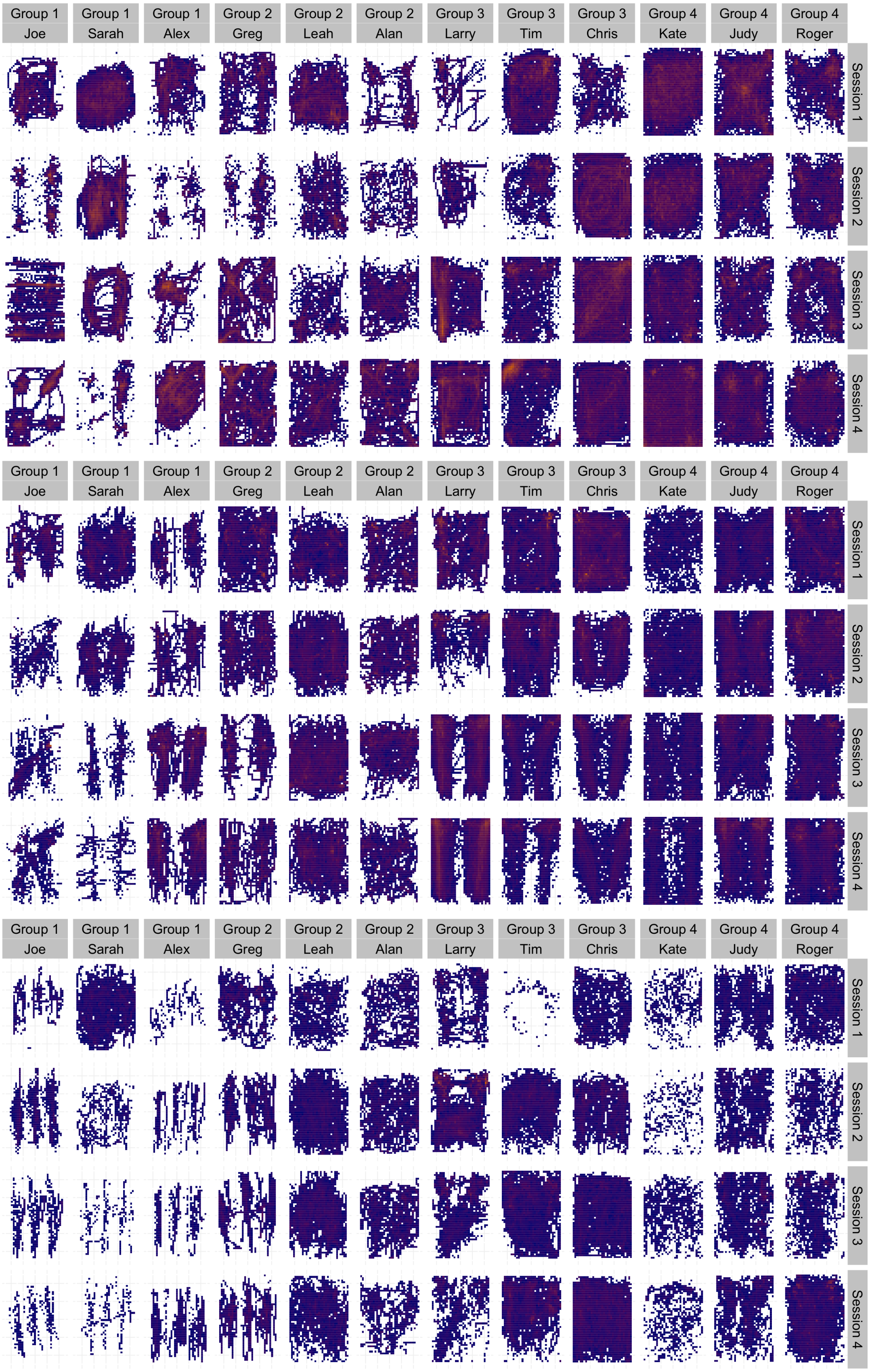


Figure 8.4.: Touch heatmaps over time, showing each musician's development from the first session (top row) to the final session (bottom row) for each of the different modes (as in fig. 8.3).

It is interesting to see how some of the distinctive features of each group emerge over the course of the four sessions. For instance, the distinctive 2-finger vertical stripes of group 2 are much more pronounced in the 3rd and 4th session than in the first two sessions. Similarly, the 1-finger 'X' shape of group 2 starts to emerge in **Greg**'s trace in session 3, and is then copied by the other musicians in the group in session 4. Group 4 evolves in the opposite direction—there are some noticeable red 'X' patterns in the earlier sessions, but they have all but disappeared by week four.

One limitation of the finger trace plots is that they present the data in a time-aggregated fashion. As I have discussed, this is helpful in visually discerning some patterns of interaction, but there is value in considering the touch traces on much shorter timescales as well.

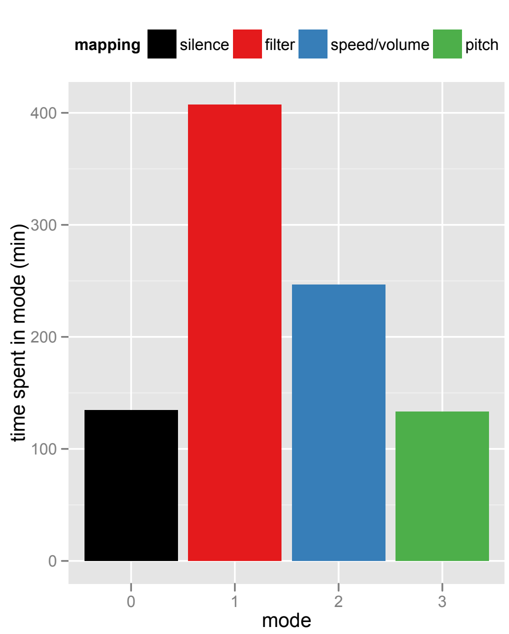
The breakdown of the time spent using the different modes (that is, different number of fingers on the screen) is shown in figs. 8.5 and 8.6. Unlike the different modes in v1 (see fig. 4.6) and v2 (see fig. 5.7), finger touches in *all* of the modes in the v3 interface had a direct influence on the sound—there were no loop period and offset parameters (as mentioned in section 6.2). So there is not the same distinction between timbral and rhythmic parameters that there was in the previous versions of the system.

The general trend in these figures is an inverse relationship between the number of fingers (the mode number) and the time spent in that mode, with silence (mode 0) about as popular as pitch shifting (mode 3). Group 1 exhibit this trend most strikingly while group 3 come closest to bucking this trend (fig. 8.5(b)). The general trend may be due to the physical affordances of the touch screen—it may have been considered more effort to use a three finger gesture than a two or one finger gesture. Alternately, this may be due to the fact that the sound mapping became (vaguely) less subtle as the number of fingers increased—filtering was subtler than volume changes which was more subtle than pitch shifting. It is hard to make this argument convincingly, though, because these figures only indicate the amount of time spent in each mode, not the parameter values or the rate of change in parameter values, which have a significant effect on how drastic the changes to the sound are. Even these statistics are misleading, because which sample (or synth) is being played has a large effect on the aesthetic impact of the signal processing operations afforded by the interface. Features of the output audio are perhaps a better way of considering this phenomenon, and will be considered in the classification process (see section 8.3).

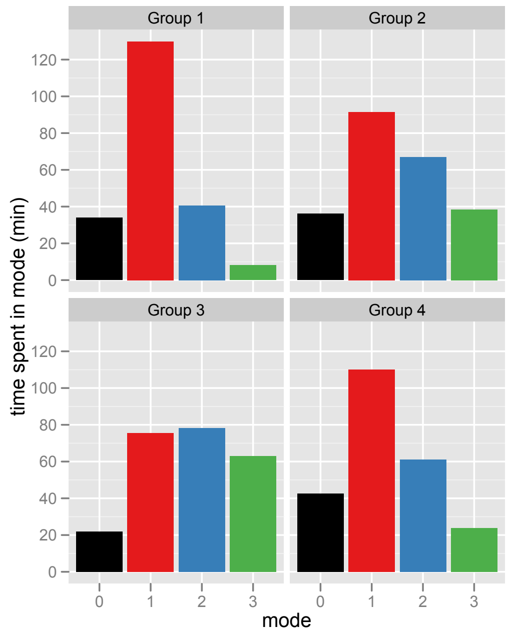
8.1.2. Sound visualisation

The finger position and mode plots are one interesting view of the data, but it is worth examining other aspects of the jam sessions as well. The finger patterns are simply a means to an end from the musician's perspective. The primary artefact of the jam is the *sound*.

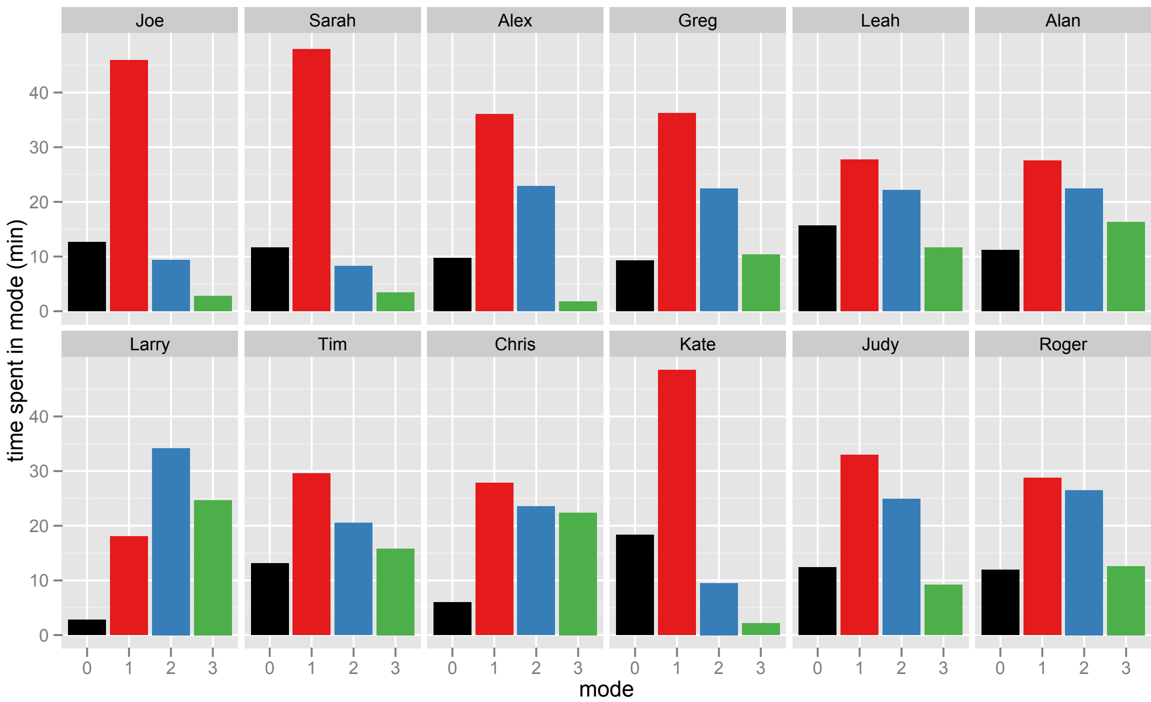
There are many ways to visualise audio data, from low-level representation of spectral power to higher-level audio features which communicate large scale changes in dynamics or timbre. The feature extraction process for audio signals is well studied, particularly in the field of Multimedia Information Retrieval (MIR) (Jain, 2008). In such studies, the overall goal is often to determine the style or instrumentation of a piece of music, or to calculate a similarity measure between two audio signals. Perhaps the most obvious audio feature to



(a) Overall mode time breakdown.



(b) Mode time breakdown by group.



(c) Mode time breakdown by musician.

Figure 8.5.: Breakdown of the total time spent in each mode. Note that mode 0 (in black) represents silence—no finger touches. More detail about the sonic effects associated with each mode is given in table 6.1.

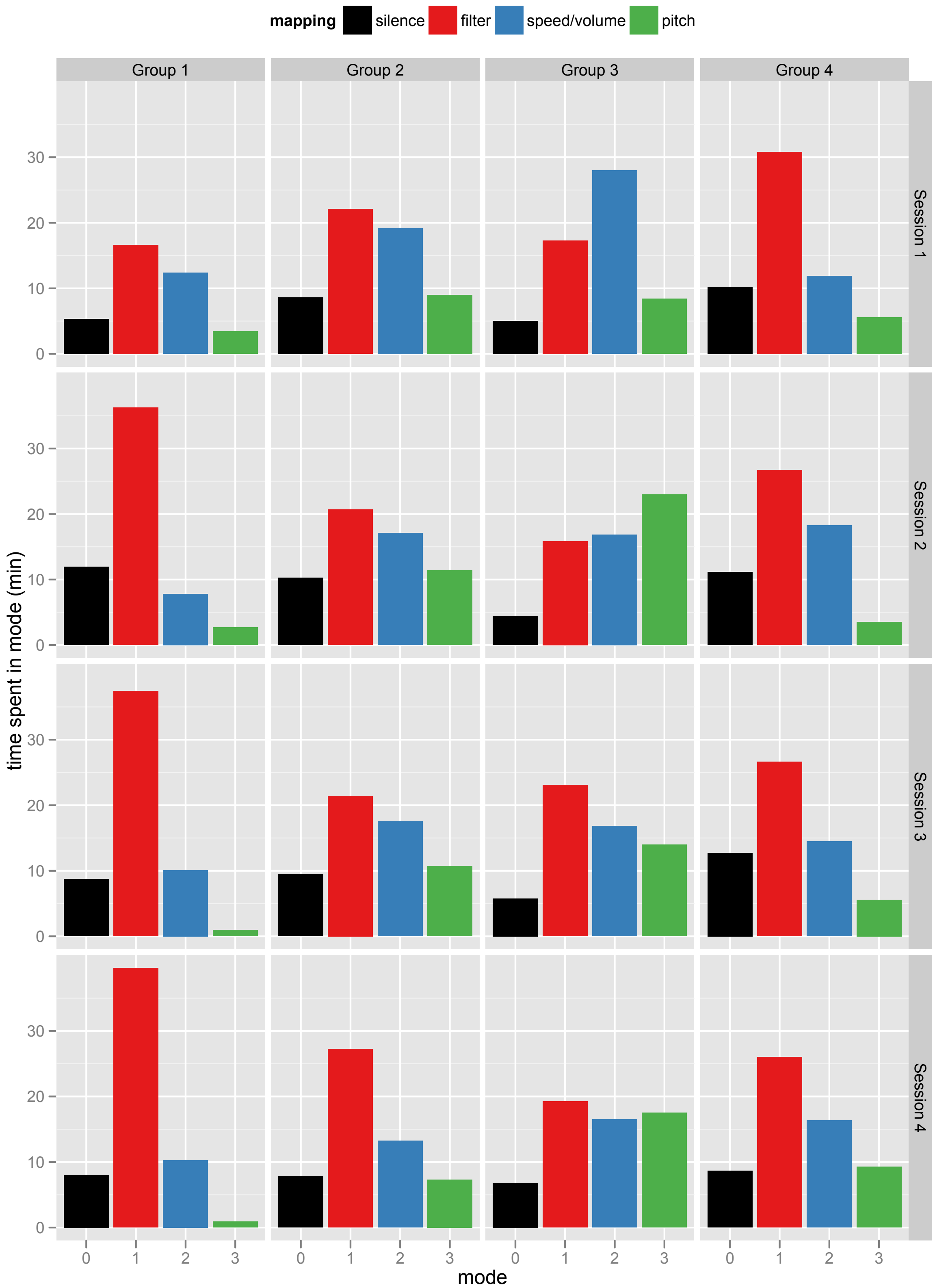


Figure 8.6.: Mode time breakdown by group and session. Higher-level views of this data (with musicians, groups and sessions aggregated together) are shown in fig. 8.5.

Figure 8.7.: Loudness (normalised, perceptually adjusted) over time. The dots represent the loudness value for each 5s window, while the line is a loess-smoothed moving average.

examine and visualise is the overall loudness of the sound made by the group.[2] The loudness data is shown in fig. 8.7. This loudness value (taken from Moore et al. (1997)) is based on a 5 second sliding window method (the same length of time suggested by Dannenberg et al. (1997)), and the signal energy is scaled to match the frequency response and hearing sensitivity of the human ear (see section 8.3 for a discussion of the choice of a 5s window). This value takes into account the nonlinearity of human auditory perception, and so is a reasonable estimate of the perceived loudness of the sound as experienced by the musicians. Also, the loudness values are normalised so that the loudest segment across all the groups corresponds to a loudness value of 1 (which happens in session 3, group 3) and the quietest sound (silence) gets a loudness value of 0.

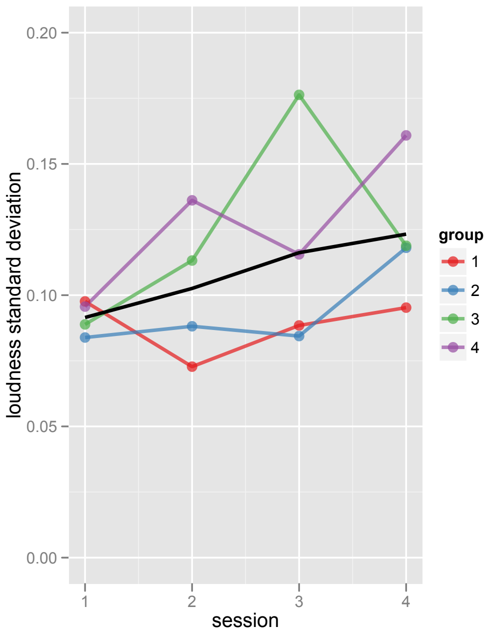
The loudness of the jams is interesting because it was dependent on the way that the musicians played. Loudness was one of the primary musical control dimensions, and, since all the musicians had the same instrument any differences in the loudness of the sessions is due to differences in the way that the different groups played their instrument. Some of the jams exhibit large dynamic shifts, from soft to loud and back again, while others were more stable. This suggests a sensitivity on the part of the musicians—they did not simply play as loud as they could all the time, despite the advantages that confers in terms of identifying their sound and the satisfaction of hearing one's own influence.

Loudness is *not* the same thing as musicality. It is often the quietest passages of music which move us the most, and good music making is characterised by both loud and soft passages. The dynamic range in the groups, measured by the standard deviation in the loudness value over the sessions, is shown in fig. 8.8(a). Looking at the overall movement (black line) in the standard deviation over the four sessions, there is a gentle upward trend, but this is not consistently shown by the groups, with group 3 peaking strongly in session 3. As discussed in section 7.2.2, this was a session where the group were exploring synthetic and digital timbres, and was a powerful example of the affective atmospheres described in the previous chapter.

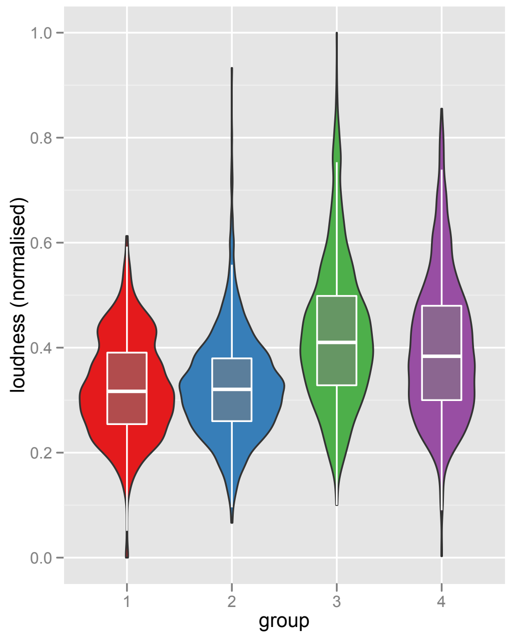
In the violin plot (fig. 8.8(b)) group 3 and group 4 are louder overall than the other two groups. Their loudness distribution appears reasonably symmetric as well. By contrast, group 1 has an obvious secondary 'bump' in density near the top of their loudness range. This seems to indicate more of a 'two-speed' loudness profile for group 1. This means that group 1 preferred either loud or soft sounds, not using the shades of grey available in between very much. The musicians did not mention anything about this in the group interviews, so it is difficult to draw any conclusions about why this may have been the case.

There are many other audio features which could be used to describe the data, although perhaps none which allow such a clear intuition for their meaning as loudness. I shall deal with alternate audio features in appendix A.6. However, it is clear that there are again real differences between the groups in the audio data, even though they were using the same instrument. This provides further support for the thesis that these differences might be useful in training a classifier to recognise the unique traits of each musician and group as they jam.

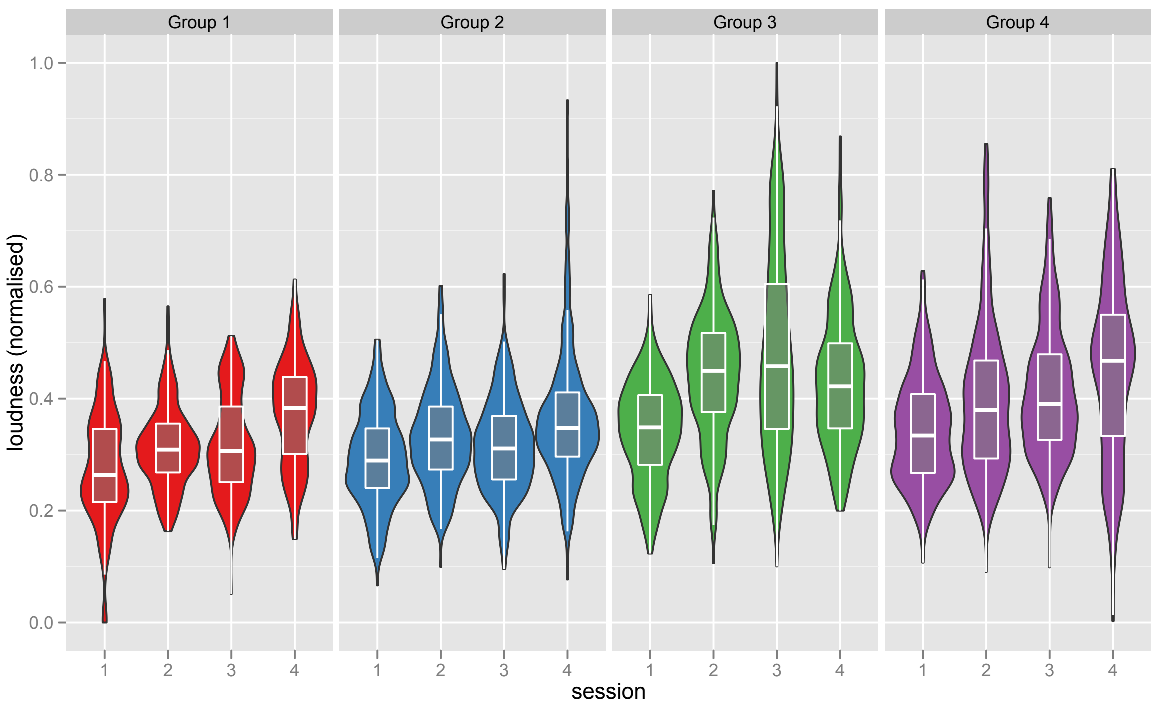
[2] Because of the way the system works (see section 4.1.1) the log contains only the audio output of each group *as a whole*, not for each individual musician.



(a) Standard deviation of the normalised loudness value over the four sessions. The black line represents the mean value across all groups. While this is not the same as the dynamic range ($\text{loud}_{max} - \text{loud}_{min}$), it is a better indication of the use of dynamic variation by the jamming musicians.



(b) Violin plot of loudness values by group. This plot shows the distribution of the loudness, as well as the 1st, 2nd and 3rd quartiles.



(c) Loudness values broken down by group and by session.

Figure 8.8.: Distribution and variation of the loudness of the group's sound.

8.1.3. Accelerometer Visualisation

The final data source I shall consider is the accelerometer data. Each iPhone had a 3-axis accelerometer, which ran constantly during the jam sessions at a sample rate of 100Hz. This accelerometer data recorded the orientation in which the device was being held (because the device is always registering the force of gravity) and also any movements (accelerations) the device underwent during the jam. While the accelerometer is often used as a control input in mobile music systems (e.g. in Camurri et al., 2010), in Viscotheque v3 this input was not mapped to any musical function. The accelerometer data, then, gives a 'surreptitious' look at the way the musicians held and used the device. Differences between the musicians in *this* data tell a different story to the finger position and audio data. While those traces represent the musicians' *conscious* creative manipulations in the act of music making, the accelerometer data shows if, in the course of this activity, there are any quirks in the way they physically held and shook the device. Figure 8.9 shows the mean accelerometer vector for each session. It is remarkable that the accelerometer vector is quite similar over the different sessions for each individual musician. Each musician had a preferred angle for holding the device, and this preference seems to be quite stable (at least over the four weeks of our experiment). This in itself raises interesting questions—why do the musicians always hold the device the same way, and what factors influence this? To see how the accelerometer vector varied during each session, Figure 8.11 shows the distribution of the x component for each musician, again broken into sessions. The y and z components show similar characteristics, and their plots have been omitted to save space.

The distributions in many cases are non-normal, and there appears to be significant bimodality (two peaks) in some of the sessions (e.g. **Sarah**, session 2 and **Tim**, session 2.). This could indicate two different iPhone-holding positions, perhaps right-handed and left-handed. As in fig. 8.9, each musician's mean x position seems to be distributed similarly across the sessions, with the exception of perhaps **Larry**, session 1 and **Alex**, session 4. There are differences between the musicians, however, and this indicates that the way (that is, the orientation) that the musicians hold the device may be a key differentiator between the musicians. I shall return to this idea in section 8.2.

8.2. Classification: machine differentiation

Visualising the finger touch, accelerometer and sound data suggests the possibility of using the differences between the musicians to differentiate them algorithmically. If the musicians are discernibly different in the way that they jam, then how would we go about noticing and examining these differences based on the touch, accelerometer and audio data they generate?

To do this requires a data analysis scheme which takes into account all of these aspects of the data in a unified way. The problem of recognising the underlying processes (or labels) behind a particular data stream is known in ML as *classification*. Given a representation of the musician's data $\mathbf{x} \in \mathbb{R}^p$, we wish to find the label $y \in \{m_1, \ldots, m_{12}\}$ corresponding to which musician the data belongs to.The classifier is a model

$$\hat{y} = f(\mathbf{x}) \tag{8.1}$$

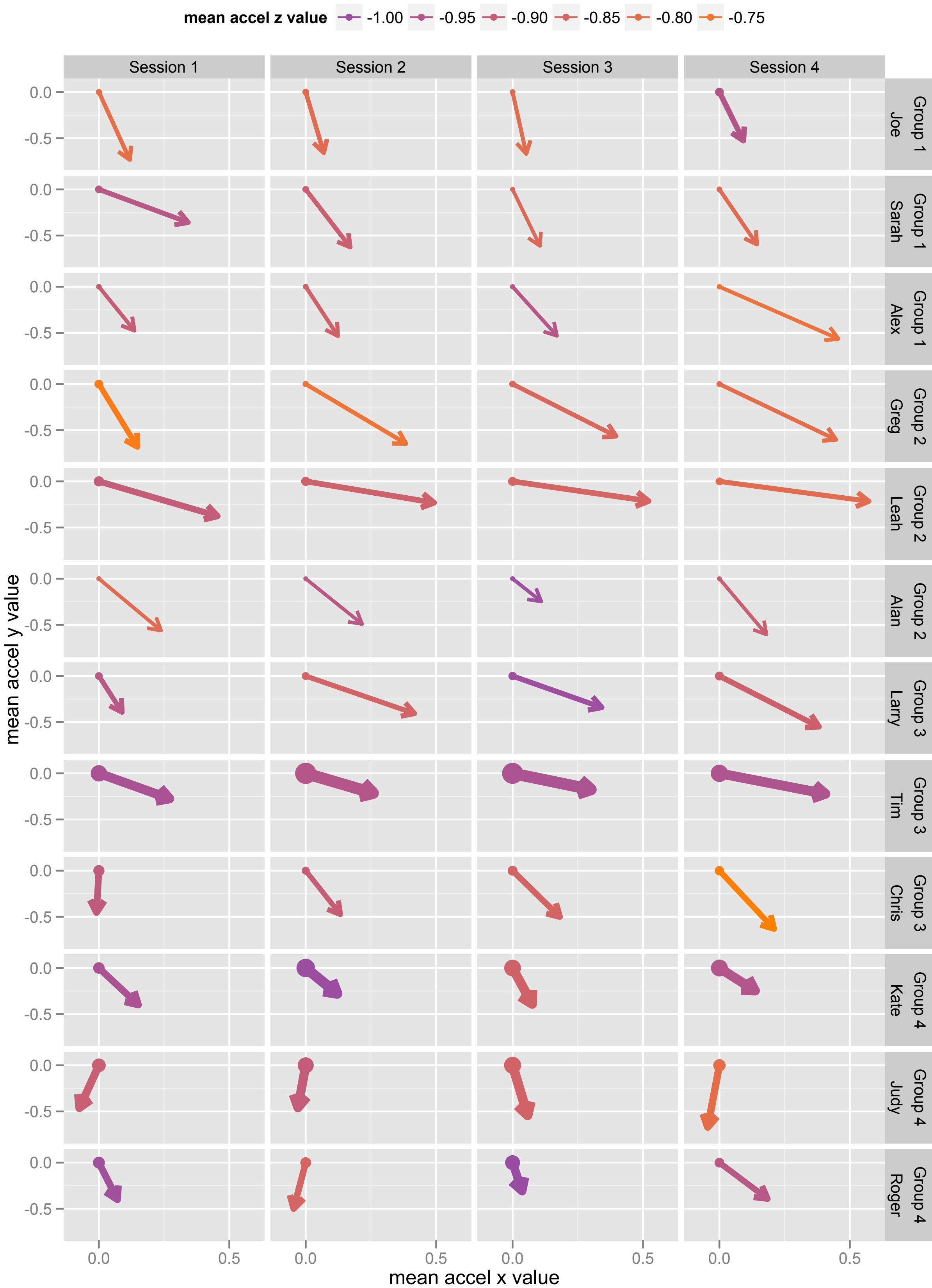


Figure 8.9.: The mean accelerometer position vector for each musician by session. The colour of the arrow shows the z component of the vector, as shown in the legend. The the thickness of the arrow represents by the mean RMS motion energy—thicker arrows mean 'more jittery' device motion. The vector points downward because the y component of the accelerometer vector is negative when the iPhone is held in the conventional 'upright' position (see fig. 8.10).

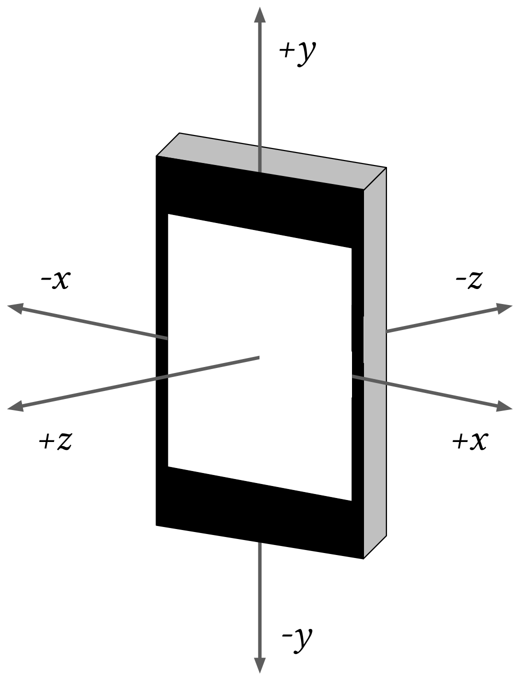


Figure 8.10.: Default accelerometer vector orientation for the iPhone.

which, given an input **x** makes a prediction $\hat{y}$ about which musician (or group) **x** comes from. By convention, the *actual* musician/group label is denoted by y, while the predicted value $\hat{y}$ is given a hat.

The goal is to minimise the Expected Prediction Error (EPE)

$$EPE = E[L(y, \hat{y})] \tag{8.2}$$

where L is a loss function, such as the simple '0–1' loss function

$$L(y, \hat{y}) = \begin{cases} 0, & \text{if } \hat{y} = y \\ 1, & \text{if } \hat{y} \neq y \end{cases} \tag{8.3}$$

which is equal to 1 if the model correctly predicts which musician the data belonged to and 0 otherwise (this function is also known as the indicator function $I(\cdot)$). This classification problem is a *supervised learning* problem: it requires a labelled training data set[3] $(\mathbf{X}, \mathbf{Y})$ such that the correct class label y_i is known for each $\mathbf{x}_i$. A supervised learning approach uses this training set to learn the associations and patterns which differentiate the different classes and aims to produce a model which can accurately predict the class of any new $\mathbf{x}_j$. There are many different algorithms available for this problem, and the best algorithm for a given problem depends on the nature of the patterns in the data and the size of the training set.

[3]note the switch to matrix notation to indicate moving beyond individual feature vectors to a data *set*—each feature vector $\mathbf{x_i}$ is a row of the matrix **X**, and the (1 column) matrix **Y** contains the real class labels for each $\mathbf{x_i}$

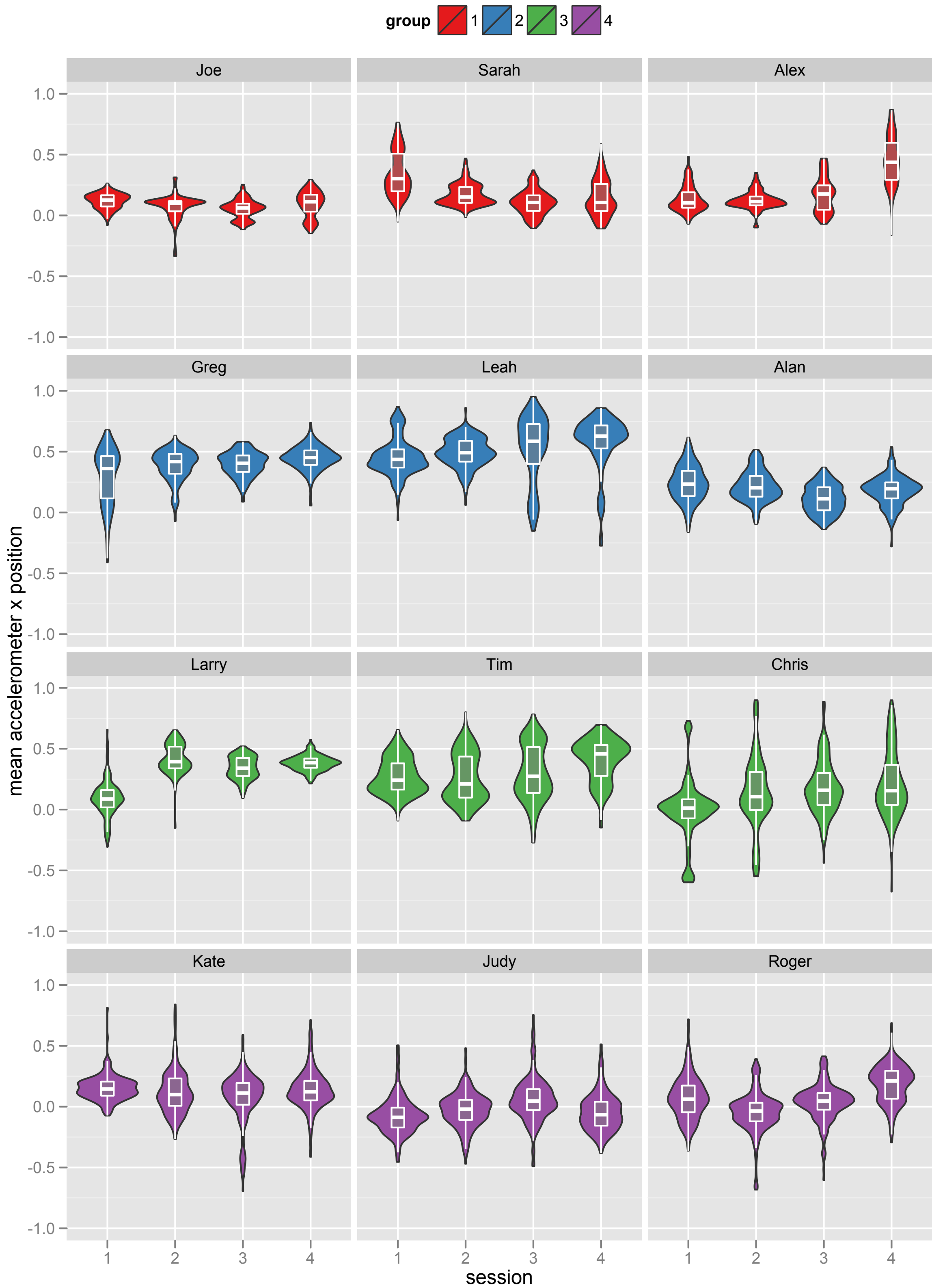


Figure 8.11.: Violin plots for accelerometer x position.

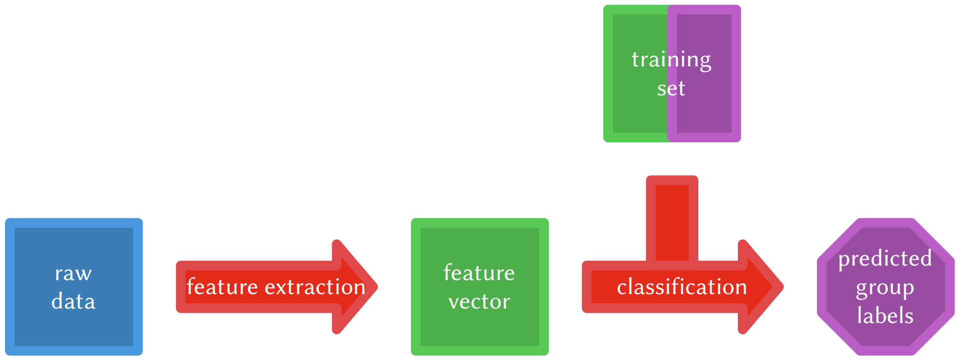


Figure 8.12.: The classification process.

In contrast, unsupervised learning describes a situation where there is no known training set—no 'ground truth'. The unsupervised learning problem, then, is to look for patterns among the data points x_i such as clusters (data points which are similar) or outliers (data points which are dissimilar to all others).

Supervised learning can benefit greatly from having standard reference data sets for a given problem. However, putting together a good quality training set is time consuming, and the 'ground truths' may have to be determined manually.

The issue of 'ground truth'—of assembling a good training set—is a pertinent one in this Viscotheque analysis. The open-ended improvisational nature of the Viscotheque jam sessions means that there is no meaningful measure of performance, and measuring the quality of each musician's subjective experience is challenging. As discussed in section 2.3, most approaches which attempt to quantify the user experience in such contexts do so with post-jam questionnaires and likert-scale responses. There is ongoing debate about the validity of such approaches in experientially-oriented activities like jamming.

To avoid the necessity of quantifying the experience, I have taken a different approach to the data analysis. I have chosen to only use groupings and response variables which arise *naturally* in the data. The most obvious example of such a grouping is 'by musician', 'by group' or 'by session'. In using these groupings in the analysis I can investigate *differences* between the musicians, which is an acceptable data-driven analogue of musical style. In doing so, I have chosen supervised learning techniques which provide insight into the structure of the problem as well as good classification accuracy. For my purposes the *model* is more interesting than classification error.

8.3. The musician as a feature vector

The general ML approach proceeds in two stages: an initial feature extraction step, followed by a classification step (as shown in fig. 8.12). The purpose of the feature extraction step is to reduce the dimensionality of the data, producing a lower-dimensional representation which still captures the salient aspects of the original data.

Choosing a set of features is an important part of the classification process. There are a few high-level tensions to manage in choosing which features to include. One issue is dealing with the temporal dimension of group music making. Music making is all about variations over time—both physically (sound waves manifest as the oscillations in the density of the medium though which they travel) and aesthetically (music which lacks any variation is boring). The feature vector representation of the jamming musician in Viscotheque should address this issue of temporality.

It is certainly possible to represent each musician's total activity (even across the multiple sessions) in a single feature vector. However, a much better approach is to use a time-windowing approach. In this approach, the musician's data is split into segments (windows) representing a time length much shorter than the length of the entire jam. A feature vector is generated for each window, using only the data that falls in that time window. Each musician's activity in a jam session is, therefore, represented by a sequence of these feature vectors.

This gives us a set of feature vectors

$$x_i, i = 1, \ldots, N \text{ with } N = MJW \tag{8.4}$$

where M is the total number of musicians, J is the total number of jams and W is the number of feature windows per jam. This feature extraction process is represented graphically in fig. 8.13.

One implication of the time-window based classification approach is that the classifier learns the patterns in the data (if there are any) that are 'stationary' over all the time windows. It is worth considering whether this is a reasonable assumption in our case. While it is true that music making is intrinsically dynamic, when we are interested in 'style' and *stable* differences between musicians, these are precisely the differences that are stable in music making. Our classifier is then primed to detect the differences that are stable. How accurately this represents the 'style' of the musicians in any meaningfully musical sense is open to debate, but the thesis of this chapter is that there are stable differences between the musicians, and my goal is to expose them to further analysis, rather than uncritically upholding them as the key to meaningful musical difference.

One other thing to consider is the *statefulness* of the Viscotheque system—the fact that the sound being produced by the device is dependent not only on the musician's finger manipulations at the current time but potentially all their manipulations up to that point in the jam. While statelessness was a consideration in the Viscotheque design (see section 6.2), a certain amount of statefulness was present in the interface. Even on a piano, use of the sustain pedal means that the sound being produced at any given moment may include notes which are ringing out but whose keys are no longer being depressed. The time window length should therefore be long enough so that an adequate amount of temporal context is included in each feature vector, while not so long as to smooth out the changes in activity which are musically meaningful and part of the natural evolution and dynamism of the jam session. A time window of five seconds is used in this analysis as a suitable compromise between these two considerations.

The Viscotheque analysis uses a 'hybrid' feature vector based on summary statistics which fall into six categories: instrument mode features, touch *activity* features, touch *position* features, touch *zone* features, accelerometer features and audio features. A complete listing of the features is given in section 8.3. Details about each feature can be found in appendix A.

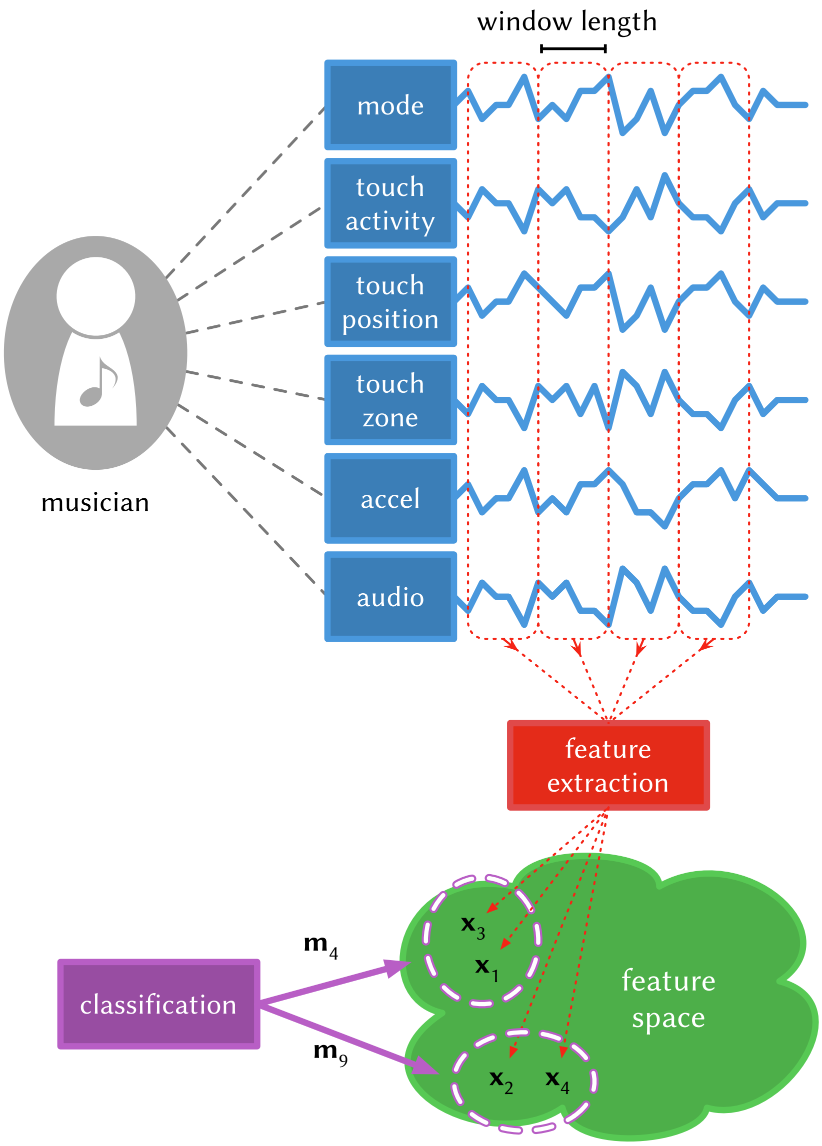


Figure 8.13.: The feature extraction process. The raw data is summarised into the different features, and then a feature vector $\mathbf{x}_i$ is generated for each feature window. The classifier then looks for patterns in the feature space and assigns a label (in this case $\mathbf{m}_k$) corresponding to the musician it predicts the $\mathbf{x}_i$ belongs to.

Feature type	Feature	Description
mode	silence	proportion of time spent silent
	mode 1	proportion of time in mode 1
	mode 2	proportion of time in mode 2
	mode 3	proportion of time in mode 3
	mode 4	proportion of time in mode 4
	sampler	proportion of time in sampler mode
	synth	proportion of time in synth mode
touch activity	touch down	number of touch onsets
	touch moved	number of finger movements
	touch distance	total distance covered by fingers
	mode changes	number of mode changes
	instrument changes	number of instrument
touch position	$\mu_{x_{\text{touch}}}$	mean touch x position
	$\mu_{y_{\text{touch}}}$	mean touch y position
	$\mu_{z_{\text{touch}}}$	mean touch z position
	$\sigma_{x_{\text{touch}}}$	s.d. in touch x position
	$\sigma_{y_{\text{touch}}}$	s.d. in touch z position
	$\sigma_{z_{\text{touch}}}$	s.d. in touch z position
touch zone	touch zone 1	see fig. A.1(b)
	touch zone 2	
	touch zone 3	
	touch zone 4	
	touch zone 5	
	touch zone 6	
accelerometer	$\mu_{x_{\text{accel}}}$	mean accelerometer x position
	$\mu_{y_{\text{accel}}}$	mean accelerometer y position
	$\mu_{z_{\text{accel}}}$	mean accelerometer z position
	$\sigma_{x_{\text{accel}}}$	s.d. in accelerometer x position
	$\sigma_{y_{\text{accel}}}$	s.d. in accelerometer z position
	$\sigma_{z_{\text{accel}}}$	s.d. in accelerometer z position
	$\text{RMS}_{\text{accel}}$	RMS energy for the accelerometer vector
audio	OBSIR 1	octave band signal intensity ratio 1
	OBSIR 2	octave band signal intensity ratio 2
	ZCR	zero crossing rate
	autoCor	audio signal autocorrelation
	loudness	perceptual loudness
	spread	spread of loudness coefficients
	sharpness	perceptual sharpness of loudness coefficients
	specFlatness	spectral flatness
	specRolloff	spectral rolloff
	specVariation	spectral variation

8.4. Building a classifier

Having extracted a feature vector from the raw data, the next step is to train the classifier. Given a feature vector $\mathbf{x}_i$ which belongs to a *musician* who belongs to a *group*, the classification problem can be formulated in two slightly different guises: with the individual musician as the classification target, or just their group.

$$f(\mathbf{x_i}) = \hat{y}_i \in \{m_1, \ldots, m_{12}\} \tag{8.5}$$
$$f(\mathbf{x_i}) = \hat{y}_i \in \{g_1, \ldots, g_4\} \tag{8.6}$$

The former shows the differences between individual musicians, while the latter only shows the differences between groups (and potentially the influence of members of each group on one another). In this section I will give results for both problems—the 'musician classification problem' and the 'group classification problem'. In describing the classification algorithms I shall primarily use the 'musician classification' notation (with target labels m_i), but the problem is equally well formed in both cases.

8.4.1. Classification algorithm

I have used three different classification algorithms: a Naive Bayes (NB) classifier, a Random Forests (RF) classifier, and a Support Vector Machine (SVM) with a radial basis function.

Naive bayes

The Naive Bayes (NB) classifier using a maximum a posteriori (MAP) decision rule and c classes is

$$\hat{y}_i = f_{\text{NB}}(\mathbf{x_i}) \tag{8.7}$$
$$= \underset{m_k}{\text{argmax}} \Pr(y_i = m_k) \prod_{j=1}^{p} \Pr(x_{ji}|y_i = m_k) \text{ for } k = 1, \ldots, c \tag{8.8}$$

with a uniform prior $\Pr(y_i = m_k) = \frac{1}{c} \forall k$ where x_{ji} is the j^{th} element of the feature vector $\mathbf{x}_i \in \mathbb{R}^p, p = 35$ and the class-wise feature distributions $\Pr(x_{ji}|y = m_k)$ are estimated (using a gaussian distribution) from the training set.

The NB classifier is so-called because it assumes that the features are independently distributed given the class label. This assumption is obviously false in our case due to the groups of similar features in the feature vector (e.g. the mean accelerometer x value and the accelerometer RMS motion energy values are not unrelated in their contribution to the class label). While this assumption is ridiculous in principle, NB classifiers often give reasonable results even when the independence assumption is violated, and they compare favourably with other algorithms in ease of implementation and training time.

The `e1071` `R` package (Meyer, 2001) provided the `R` implementation of the algorithm.

Random forests

A more sophisticated classifier is a Random Forests (RF) (Breiman, 2001). An adaption and extension of standard decision trees (Hastie et al., 2009, Ch. 9.), the RF model trains a large number (a 'forest') of T decision trees, with each tree trained on a randomly selected subset of the features in $\mathbf{x}_i$:

> A random forest is a classifier consisting of a collection of tree-structured classifiers $h(\mathbf{x}, \Theta_t), t = 1, \ldots, T$ where the Θ_t are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\mathbf{x}$ (Breiman, 2001, definition 1.1).

The RF classifier works by assigning the group label with the largest vote across the predictions of all the trees in the forest:

$$\hat{y}_i = f_{\text{RF}}(\mathbf{x_i}) \tag{8.9}$$

$$= \underset{m_k}{\text{argmax}} \sum_{t=1}^{T} I(h(\mathbf{x}, \Theta_t) = m_k) \text{ for } k = 1, \ldots, c \tag{8.10}$$

where $I(\cdot)$ is the indicator function. In his original paper Breiman (ibid.) proves results which show that the algorithm is resistant to over-fitting, and invariant under strictly-monotonic transformations of the input feature space. This second characteristic in particular is helpful in light of our 'hybrid' feature vector, where the groups (categories) of feature vectors may be similarly scaled, but inter-group scales may be wildly different.

One other nice feature of the random forests model is that it provides a feature importance measure (Genuer and Poggi, 2010). I shall make use of these feature importance results in section 8.6.1.

The `randomForest` R package (Liaw, 2002) provided the R implementation of the algorithm. I used $T = 2000$ trees and $\lfloor \sqrt{p} \rfloor = \lfloor \sqrt{35} \rfloor = 5$ randomly selected features at each node, as suggested by Breiman (2002).

Support vector machine

The final classifier model used was a Support Vector Machine (SVM). These classifiers are known for their high classification performance as well as their 'black-box' inscrutability (Suykens, 2001). SVMs work by using a kernel mapping $K(\mathbf{x}, \mathbf{x}') = \langle h(\mathbf{x}), h(\mathbf{x}') \rangle$ to calculate the inner product between input feature vectors in a higher dimensional space, then finding an optimal hyperplane in this space which separates the feature vectors of different classes. The mathematics of SVMs is complex, a good overview can be found in Hastie et al. (2009, ch 12).

The `kernlab` R package (Zeileis et al., 2004) provided the `R` implementation of the algorithm. A multi-class SVM with a radial basis function was used, and the optimal value for the hyperparameter sigma was estimated from the data as per Caputo et al. (2002).

8.4.2. Cross-validation

The training set classification accuracy for each of the models described is easy to calculate—train the model on the data, then see which musicians get classified accurately using eq. (8.3). However, this can lead to over-fitting; the situation where the model gets great accuracy on the training set but does not generalise well to new, unseen data. To understand this, it is important to realise that what we are doing in this classification process is trying to understand the distribution of each musician's feature vector $\mathbf{x}_i \in \mathbb{R}^p$ as a p-dimensional random variable. The data collected in the Viscotheque jam sessions represents a sample (in fact, many samples—one for each time window) from this distribution, but the true distribution is unknown. Again, this is the stationary distribution which represents those aspects of music making which do not change moment-to-moment (a musician's 'style') rather than the natural variations which are a key feature of their musical trajectories in the jam session.

There is a trade-off between bias (how closely the model is expected to get to the true value) and variance (how susceptible the model is to variations in the sample used for training) in model fitting. If the selected model fits the data too well (low bias), the model will be too specific to the sample used for training, and will perform poorly on a new sample from the same distribution (high variance).

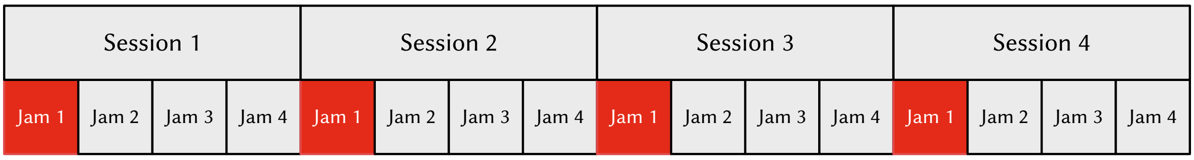
The usual approach to this problem is to use Cross-validation (CV). CV works by partitioning the data into a training set and a test set. The model is trained on the data in the training set alone, and the Expected Prediction Error (EPE) is obtained by seeing how it performs on the test set data. Often, n-fold cross validation is used, where the data is partitioned into n partitions and the EPE calculated by taking the average prediction error over all possible combinations of $n - 1$ partitions for the training set.

In this analysis, I have used a CV approach which takes into account the natural divisions in the data. Instead of randomly partitioning the data into training and test sets, I use the *jam* and *session*[4] partitions to naturally break the data into training and test sets (see fig. 8.14). Using the jam-based partitions for our test set (as in fig. 8.14(a)), the test set contains data from a (5 minute) *jam* which was unseen by the model in the training phase. The EPE of the model is therefore an indicator of differences in the data which persist from jam to jam. Perhaps even more interestingly, when using the session-based partitions the test set is made up of feature vectors which were from a whole new *session*—a new week of jamming. The EPE in this case is then an indicator of differences which persist from week to week.

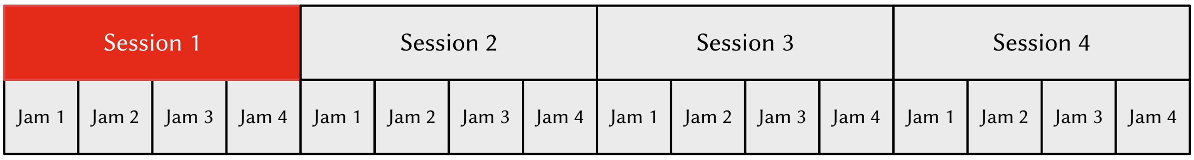
8.5. Classification results

Using R, the EPE was estimated for each of the classifiers as described in section 8.4.1. Two different 'targets' were used—classifying on musician (12 class labels), and on their group (4 class labels). Also, two different CV regimes were used as described in section 8.4.2, both jam-based and session-based CV. The EPE results are shown in fig. 8.15 and table 8.1. The EPE is shown in fig. 8.15(a), while the same error rate expressed as a multiple of chance (that is, the error rate of assigning a label at random) in fig. 8.15(b). This value differs depending

[4]Remember that each weekly *session* was composed of four 5 minute *jams*.



(a) Cross-validation by jam.



(b) Cross-validation by session.

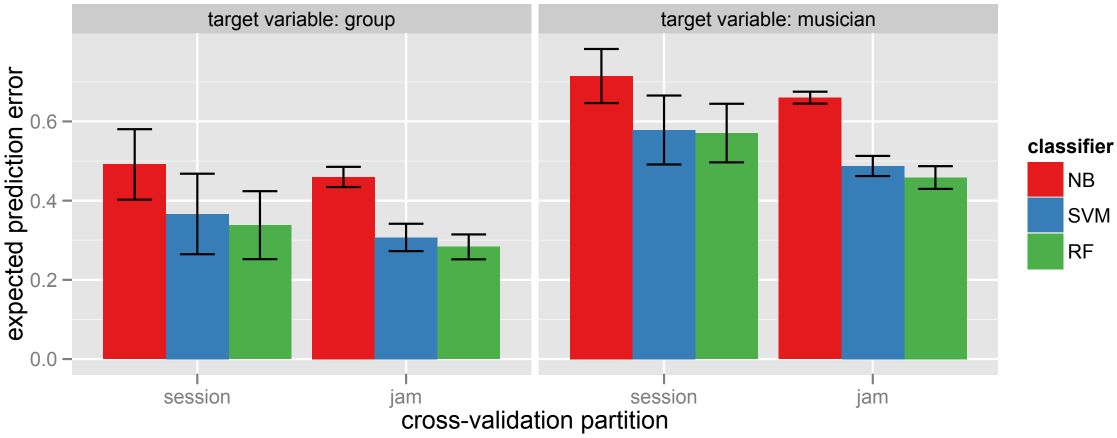
Figure 8.14.: The cross-validation technique. In these figures the first jam (fig. 8.14(a)) or session (fig. 8.14(b)) is used as the test set (shown in red), and the rest of the data is used as the training set. The expected prediction error is obtained by taking the mean of the test set error rate over all 4 such partitions.

on whether the musician or the group was the classification target because of the different number of potential labels in each case. In each case, the NB model performs significantly worse than the other two models. The RF and SVM models are indistinguishable within their error ranges, with the RF model performing *slightly* (but not significantly) better in each case. The raw EPE is better when classifying by group rather than by musician, but when the different number of class labels is taken into consideration (as in fig. 8.15(b)), the classifiers perform better when classifying by musician rather than by group.

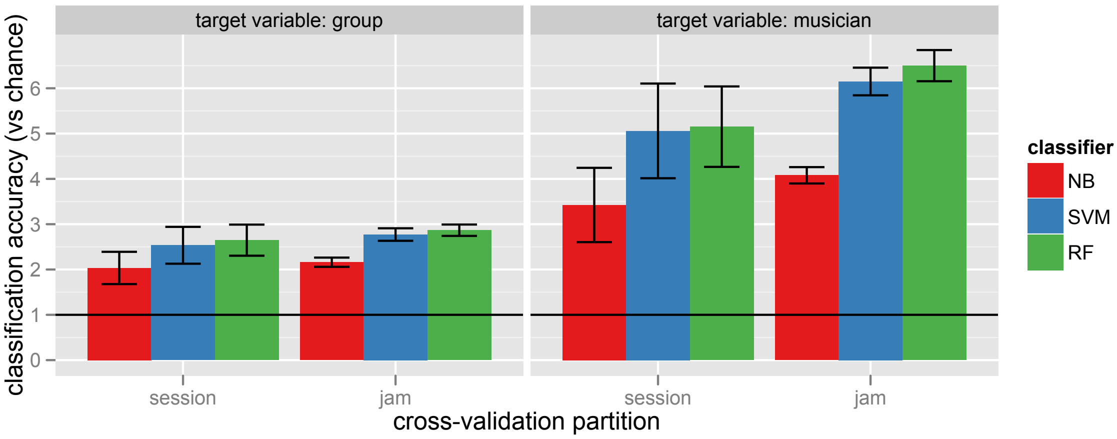
When cross-validating with the leave-one-session-out method, the EPE achieved by the RF and SVM classifiers is around 0.6 (60%) on musician and around 0.35 (35%) on group. This means that based on just 5 seconds of log data from a given musician, having not seen any other data from that musician for that session, the RF or SVM classifiers are 5 times better than chance at guessing which musician the data came from and 2.5 better than chance at guessing which group they came from.

When cross-validating using the leave-one-jam-out method, the RF and SVM classifier EPE is below 0.5 (50%) on musician and around 0.3 (30%) on group. This means that based on just 5 seconds of log data from a given musician, having not seen any other data from that jam (but having seen data from the other jams in that session), the RF or SVM classifiers are 6.5 times better than chance at guessing which musician the data came from and almost 3 times better at guessing which group the data came from. These results are better than the CV on session results. This is unsurprising, as one would expect more consistency between musicians over the different jams *in the same session* than one would expect week-to-week.

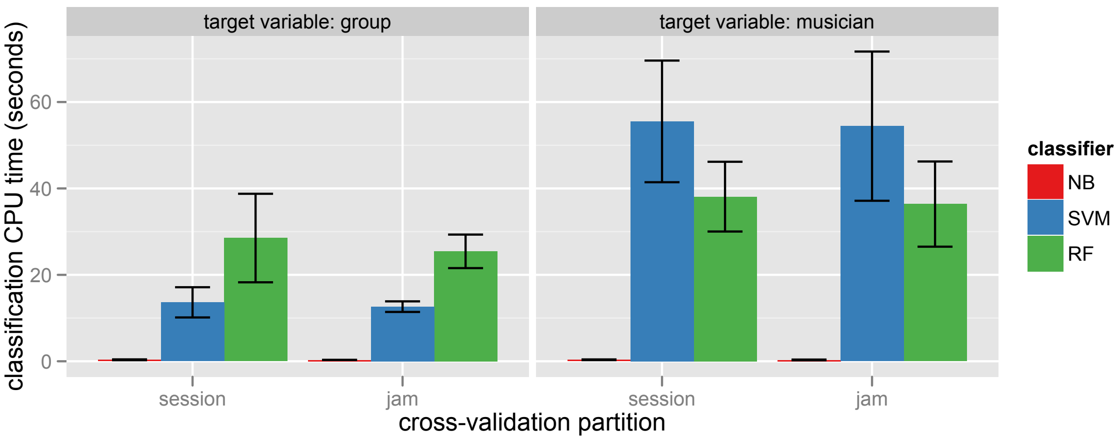
Another aspect of the classification process is the CPU time required to fit the model on the training set and test the predictions on the test set (see fig. 8.15(c)). The simplicity of the NB classifier leads to an obvious performance win over the other two classifiers. The SVM appears more susceptible to the cardinality (size) of the set of target labels—with significantly worse performance (about twice as bad) when classifying on musician (12 labels) vs classifying on group (only 4 labels). The RF classifier CPU time is more consistent between the two



(a) Classification EPE (lower is better).



(b) Classification accuracy, expressed as a multiple of chance (higher is better).



(c) Classifier CPU time (lower is better).

Figure 8.15.: Classification results for different classifiers based on 5s long feature windows. Error bars represent the s.d. in prediction error across the four cross-validation folds. The CPU time required to train the model is shown in fig. 8.15(c).

Target	CV by	Classifier	EPE	PE s.d.	vs chance	CPU time
group	session	NB	0.49	0.09	2.0	0.34
		RF	0.34	0.09	2.6	28.52
		SVM	0.37	0.10	2.5	13.64
	jam	NB	0.46	0.03	2.2	0.30
		RF	0.28	0.03	2.9	25.44
		SVM	0.31	0.03	2.8	12.64
musician	session	NB	0.71	0.07	3.4	0.34
		RF	0.57	0.07	5.2	38.10
		SVM	0.58	0.09	5.1	55.52
	jam	NB	0.66	0.02	4.1	0.31
		RF	0.46	0.03	6.5	36.38
		SVM	0.49	0.03	6.1	54.41

Table 8.1.: Expected value and standard deviation for prediction error, by model and CV scheme (see fig. 8.15 for graphical representation).

classification targets, with slightly worse performance on the smaller ‘group’ target.

One way to determine whether these models are detecting *real* structure in the data (and therefore real differences between the musicians) is to run the classifier on appropriate synthetic data. Figure 8.16 shows the results of this process using a random permutation of the rows of the data matrix **X**. This permutation has the effect of jumbling up the class labels in the training set. The overall distribution of each feature (that is, each column of **X**) remains the same, only the structure attributable to the differences between the classes (that is, between the musicians) is lost. As can be seen in fig. 8.16, the classification accuracy falls to chance levels as expected. This is a good indicator that the differences between the musicians are real, and not artefacts of the classification process. It is hard to know what to make of these results—are they *good?* The Viscotheque is a custom-made music making system, there are no other research teams using it to compare our data against. There are similar iPhone-based group music-making systems, but this kind of holistic ‘musician recognition’ is rarely attempted. So how can we assess the quality of these results?

It is worth making two points in response to this question. Firstly, the fact that the classifiers perform significantly better than chance indicates that there is real structure in the data. As lossy and reductionistic as it may seem to represent each musician as a sequence of feature vectors in $\mathbb{R}^p$, the distribution of these vectors does differ between the musicians. Because each musician used the exact same instrument, over the exact same number of sessions, any differences between them must be due to either their own intrinsic style and sensibilities or the influence of the other members of the group. Although the log data does not tell the whole story (important things like facial expressions or each musician’s mood on a given day cannot be ascertained from the touch and accelerometer data), the patterns which the

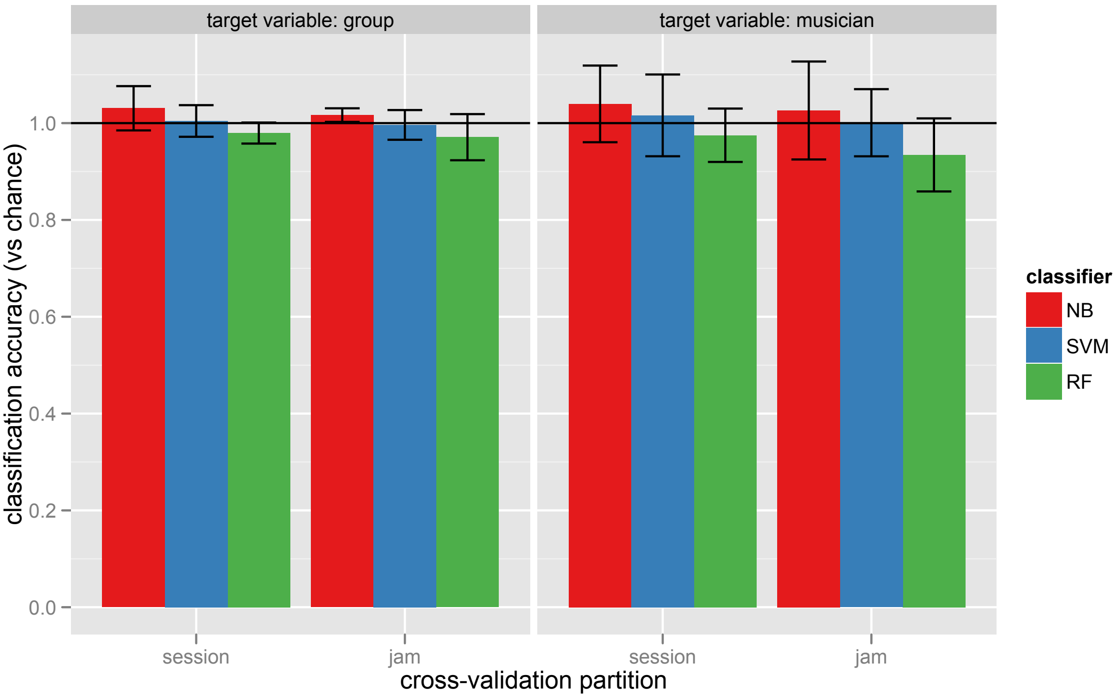


Figure 8.16.: Classification accuracy (times chance) for 'synthetic' jumbled up data. As expected, the classifiers all perform at around chance, i.e. they detect no salient differences between the synthetic musicians.

data does exhibit are worth closer inspection.

This leads to the second point about assessing the quality of the modelling and classification results. The purpose of the classification process is not classification accuracy for its own sake. We are interested in what these models can tell us about *why* the musicians and groups are different. Some models are better than other in this regard, and in the next section I shall drill down into the RF model in particular to examine the nature of the structure in the data and the differences (and similarities) between musicians and groups.

8.6. The differences in detail

In this section I shall present insights from the RF model into the nature of these differences. The primary reason for favouring the RF model in the rest of this analysis is its interpretability, as discussed in section 8.4.1.

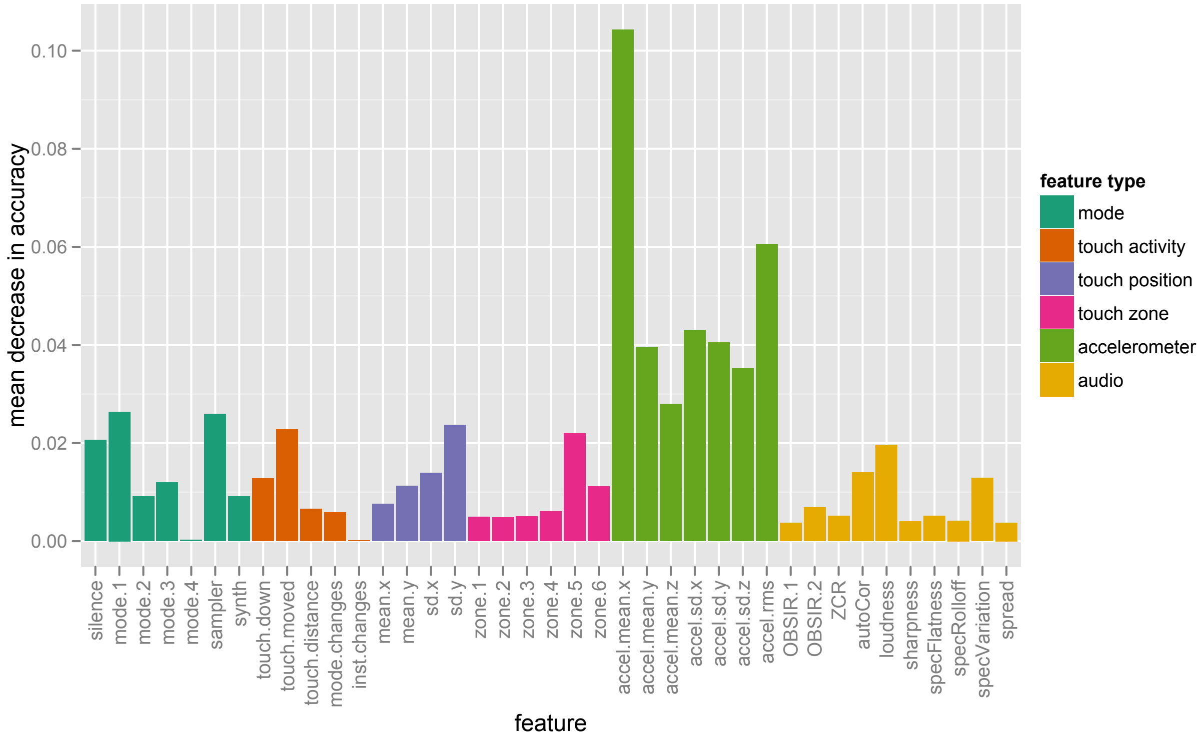
8.6.1. Feature importance

Because each tree in the forest is trained on a different subset of the features (that is, a different subspace of the input feature space), a measure of the importance of different features can be obtained by looking at the effect that including or excluding a particular feature has on the accuracy of the classifier. The RF model allows us to calculate the mean decrease in accuracy due to *excluding* a given feature. This quantity is an indicator of how important each feature is in differentiating between the class labels. A higher value (that is, a larger mean decrease in accuracy) indicates a more important feature, and vice-versa. The mean decrease in accuracy is shown in fig. 8.17 for both classification by name and by group. Interestingly, the accelerometer features are the most important features from the perspective of the RF classifier. In particular the mean accelerometer x value is the most important feature in both the musician and group classification tasks. This axis represents the 'roll' of the device (in the pitch-roll-yaw sense, see fig. 8.10, which may indicate whether the device is being held in the right or left hand as well as the preferred angle for holding. The classifier accuracy broken down by feature type is shown in fig. 8.18.

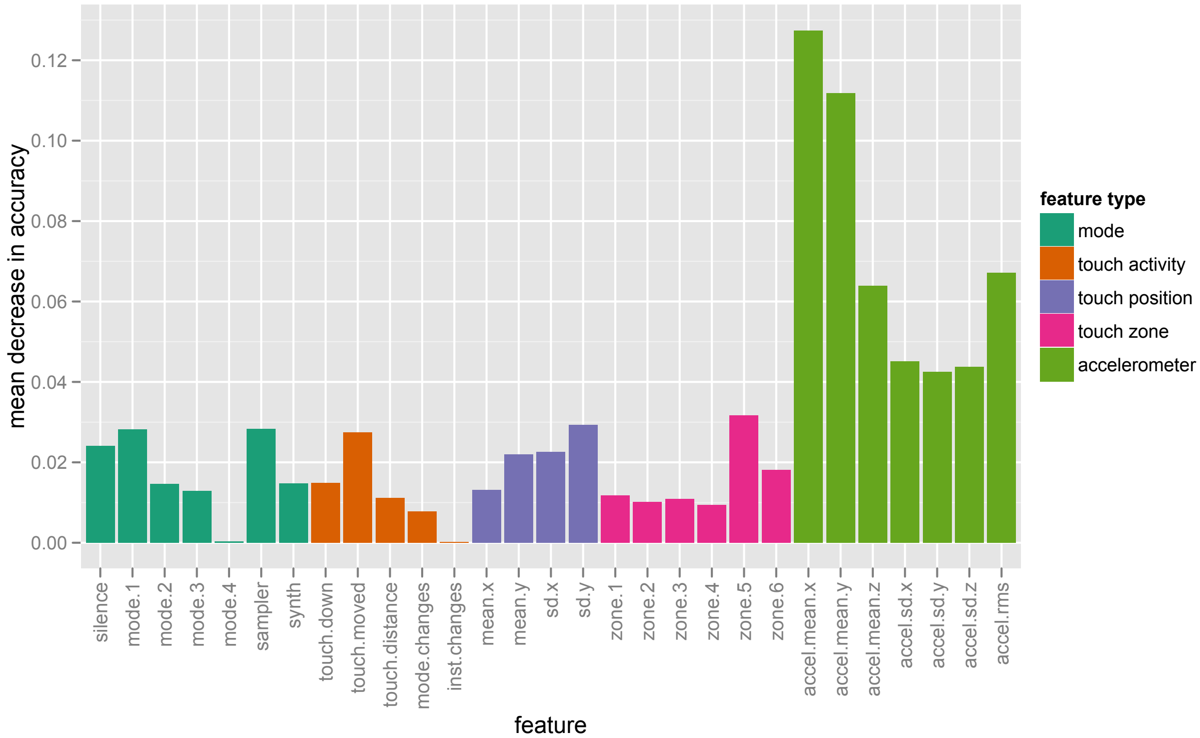
Looking back at fig. 8.9, the differences in mean device orientation between musicians are clear, and it is unsurprising that the classifier picks up on these features to discern between the musicians. Again, the consistency of these features across sessions is interesting, and the RF importance results reinforce that the way in which each musician holds their iPhone is not determined by how they are feeling on the day or where they are sitting, it is an element of their *style* in playing the instrument (if style is used to denote anything about their behaviour which differentiates them from another performing the same activity).

A couple of other aspects of the feature importance are worth noting.

- The most important feature outside of the accelerometer features is the 'touch moved' feature. This feature indicates how often any finger was moved on the screen, regardless of the speed or location of that movement.



(a) Feature importance for classification by group.



(b) Feature importance for classification by name.

Figure 8.17.: Feature importance as measured by mean decrease in accuracy when the feature is left out of the decision trees. The audio features are not included in fig. 8.17(b) because the audio data is the same between the musicians in the same group.

- The most important audio feature was the loudness (see fig. 8.7 for a visualisation of the loudness time profile and appendix A.6 for a description of the feature). This is also the most meaningful feature from an explanatory standpoint—it is easy to form an intuition about what it means for music to be loud, but not so easy in the case of, say, spectral variation.

- The most meaningful touch zone feature was the top left-hand corner (zone 5), followed by the top right-had corner (zone 6). This suggests that the musicians differ most in their usage of the top of the screens. Whether this is for musical reasons or due to the physiology of the hand or the influence of the visuals is not clear.

The importance of the accelerometer data suggests some interesting possibilities for future research. If the differences between musicians in the way that they hold the device are as distinctive and stable as the Viscotheque v3 jam data suggests, then this data could be used to identify musicians without them having to manually identify themselves via the interface (as was the case with the Viscotheque instrument). This could be used for convenience, reducing by one the (already small) number of steps required to jam with the Viscotheque DMI. Alternatively, this may prove useful in a scenario where the device is passed around amongst a group of musicians, such as in a group context where there are not enough devices to go around. In this case, the accelerometer data could be used to determine when the device had changed hands and update the state of the system accordingly.

This result also suggests more nefarious uses, such as surreptitiously tracking users via a sensory modality which they are not expecting to be identified by, and therefore are not careful to guard against. This is particularly salient because the accelerometer data was not *doing* anything musically in the jam sessions—it was not part of the sound mapping. Instead it represents a measure of the user's activity which they were *not* consciously manipulating, and is as a result perhaps even a more useful for surveillance purposes. There are lots of open questions surrounding the potential effectiveness of accelerometer tracking from this perspective, but the Viscotheque v3 jam sessions suggest at least that there are hidden 'signatures' in the way that individuals hold and move their device while engaging in the activity of music making.

Finally, this result is interesting inasmuch as it was a surprise in the data which was not picked up in the interviews or expert observations of the jamming group. Because the experiment was not set up to examine this phenomenon then it is very difficult to determine the cause or the contributing factors, but it does suggest possible future experiments which are tailored to answering those sorts of questions.

8.6.2. Class confusion

Aside from the feature importance measures, the other insights to be gleaned from the models comes from the *confusion matrix*, which indicates which musician/group labels are most often confused. In this section I shall use the confusion matrices from the RF model. I have trained the model on the complete data set in in this case, since the out-of-bag error rates[5] provided

[5]The out-of-bag error rate is calculated by testing the prediction accuracy where each input vector $\mathbf{x} - i$ is classified using only trees in which it was not part of the training sample used to grow the tree

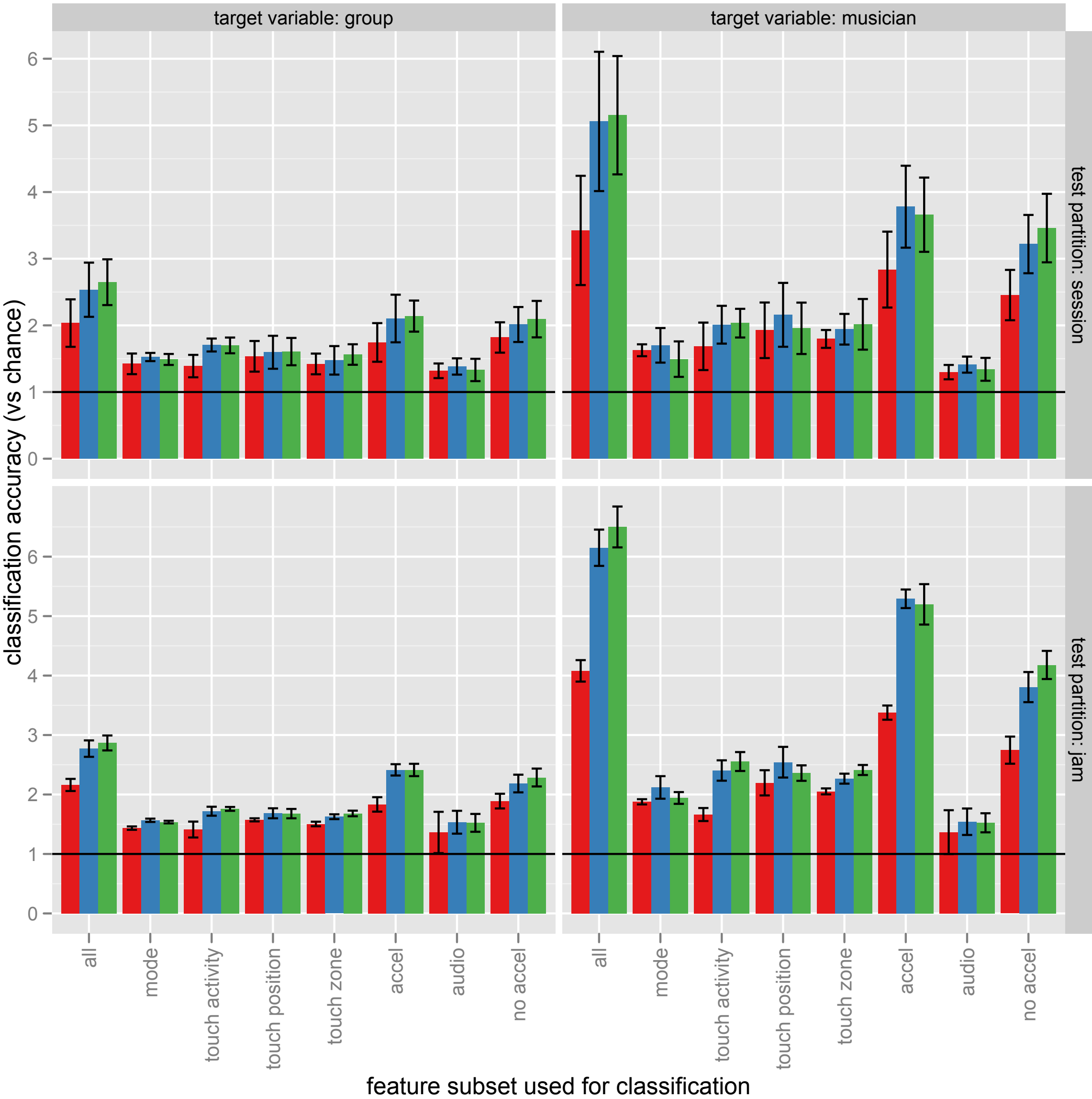


Figure 8.18.: Classification accuracy (vs chance) for all of the different feature types. Classification results are shown for each figure type in isolation, as well as for all the features together and all the features *except* the accel features. In each case, the accuracy is greater when using the accel features alone than in the case where *all* the other features are used.

	group 1	group 2	group 3	group 4	EPE
group 1	2195	234	77	111	0.1613
group 2	235	2283	186	134	0.1956
group 3	77	182	2162	459	0.2493
group 4	128	130	285	2316	0.1899

Table 8.2.: RF group confusion matrix. A visualisation of this confusion matrix is shown in fig. 8.19(d).

	Joe	Sarah	Alex	Greg	Leah	Alan	Larry	Tim	Chris	Kate	Judy	Roger	EPE
Joe	698	66	21	2	0	27	0	3	21	9	12	16	0.2023
Sarah	73	542	46	37	31	39	13	8	28	29	15	13	0.3799
Alex	58	76	505	56	16	73	15	8	18	16	9	18	0.4182
Greg	8	24	33	792	6	17	23	0	17	2	16	8	0.1628
Leah	2	4	11	52	728	33	22	37	17	20	8	12	0.2304
Alan	20	51	59	39	14	650	20	4	37	13	8	31	0.3129
Larry	4	4	18	39	18	11	725	51	21	28	15	26	0.2448
Tim	1	2	6	6	52	7	42	710	18	58	26	32	0.2604
Chris	15	18	12	9	41	40	32	27	572	31	118	45	0.4042
Kate	45	21	14	4	15	24	4	46	33	615	79	53	0.3547
Judy	15	6	4	8	4	11	15	22	56	61	670	81	0.2970
Roger	19	11	5	15	12	37	42	34	50	85	106	537	0.4365

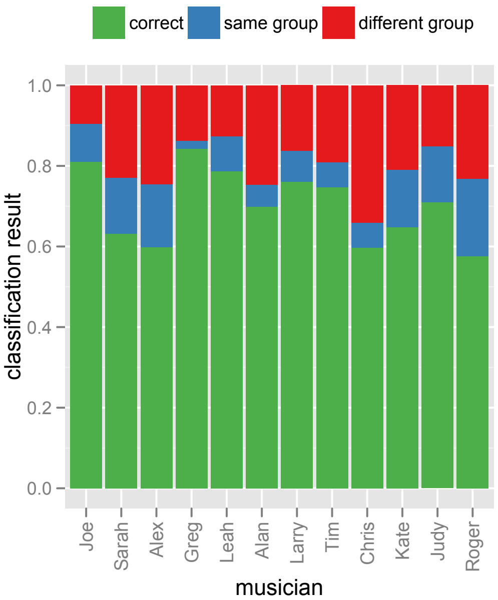
Table 8.3.: RF musician confusion matrix. A visualisation of this confusion matrix is shown in fig. 8.19(b)

by the model are an unbiased estimate of the true error rate (Breiman, 2001).

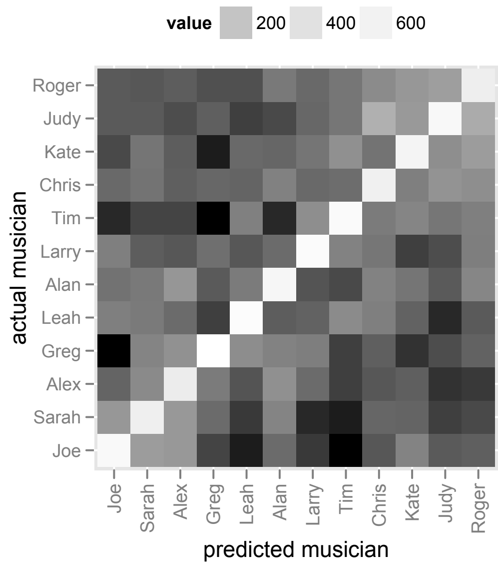
The confusion matrix for the classification task is shown in table 8.3 (target: musician) and table 8.2 (target: group). Visual representations of this data are shown in fig. 8.19.

In classifying by musician there does not seem to be any global pattern in 'same group' and 'different group' errors. This is unsurprising—in this classification the task the model does not know which musicians belong to which group, and so is not 'looking for' similarities and differences between groups as such. It is interesting that musician **Greg** is almost never mistaken for any other members of his group—he has a much smaller 'same group' error rate than any of the other musicians in any group.

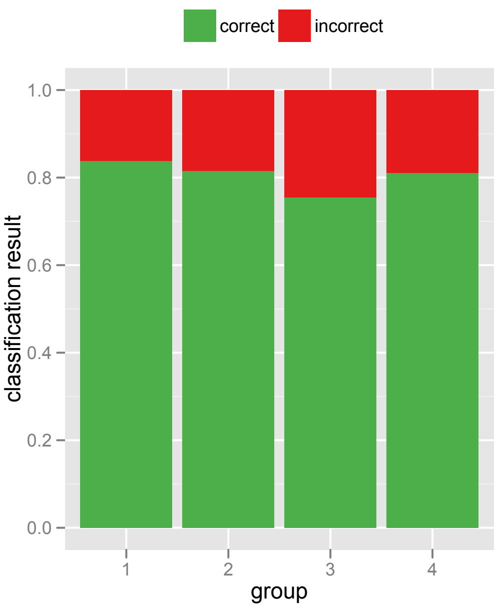
One interesting feature of the group classification matrix is the on-axis 2×2 blocks seem to be lighter than the off-axis 2×2 blocks. This means that groups 1 and 2 are more likely to be mistaken for one another, and likewise for groups 3 and 4. That these pairings (1 & 2 and 3 & 4) are contiguous is co-incidental—there was no larger pattern in the way the groups were numbered.



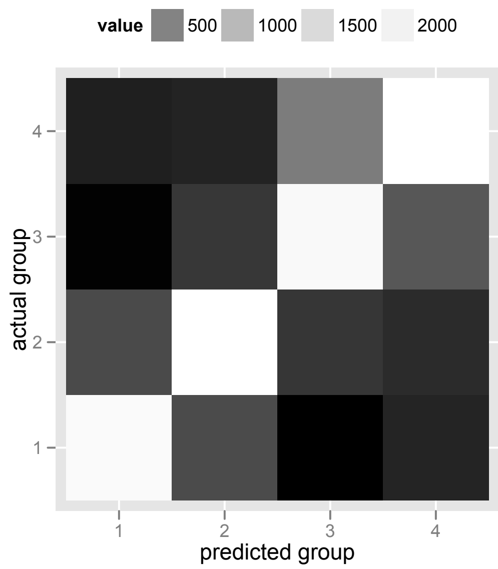
(a) Musician classification error breakdown. Errors are broken down into 'same group' (i.e. the incorrect label was from another musician in the same group) and 'different group'.



(b) Musician classification error matrix. Lighter (whiter) grid squares represent a greater number of predictions. On-axis squares represent correct predictions, while off-axis squares represent misclassifications. The raw data can be seen in table 8.3.



(c) Group classification error breakdown.



(d) Group classification error matrix. The raw data can be seen in table 8.2.

Figure 8.19.: Detailed error breakdowns for the RF classifier. The raw data for these figures can be found in tables 8.2 and 8.3

8.7. Chapter summary

In the move towards rich description and qualitative methods and the rejection of quantitative measures of experience, third wave HCI often neglects the ease with which digital interactions can be recorded and analysed. In this chapter I have attempted to examine the data in a way which avoids these pitfalls.

Visualising the touch, accelerometer and audio data from the Viscotheque v3 sessions reveals some differences between musicians and also some consistency week-to-week. By constructing a feature vector representation of this data (with a sliding 5 second time window), a Random Forests (RF) classifier was able to accurately classify which group/musician the data was generated by with accuracy of up to 6.5 times better than chance. The accelerometer data (device orientation and movement) is the most significant in differentiating between the musicians. While more work needs to be done to determine the cause of this phenomenon, it is an indicator that machine learning techniques can provide interesting insights without resorting to subjective measures of felt experience.

9. Affective atmospheres, machine learning and third-wave HCI

Jamming is an example of the type of open-ended experience-focused group activity which is associated with third wave HCI. This chapter attempts to draw together some of the themes and data analysis from the three phases of the Viscotheque design process.

The chapter begins by discussing the evolution of the sound mapping in the Viscotheque DMI through its iterative, participatory design process. This section focuses on the qualitative results of the v3 jam sessions in particular. The next section readdresses some of the theoretical tensions and disagreements in third wave HCI discourse, particularly around the notion of difference. Finally, by advocating the use of numerical techniques without resorting to arbitrary and subjective measures of experience, the chapter sets forth an agenda for taking advantage of the rich data collection afforded by the digitisation of group creative practices such as jamming.

9.1. The affective power of the ephemeral artefact

The sound mapping is the core component of a DMI. It defines how the instrument works, the way that it mediates between the musician and their sound. The main element of the Viscotheque design process was therefore the evolution of the sound mapping. Over the three versions of the Viscotheque DMI I can observe two key trends:

- towards a *natural* control of the sound
- towards a *synthetic* (digital) aesthetic in the sound

I am using natural in the sense used by Blaine and Fels (2003b) to mean a mapping which is responsive—that every action has an immediate (although potentially subtle) effect.

The trend towards a natural mapping is to be found in the move away from a 'process control' interface, where the musicians had limited control over a few key parameters of their music making, to an interface where they could trigger and morph their sound directly. The initial decision to start with a simple loop-based interface was prompted by a desire to avoid chaos (see section 4.2), but the experience and feedback of the musicians in the v1 and v2 jam sessions led to the loop-based functionality being dropped in favour of direct sample triggering in Viscotheque v3.

Similarly, in v1 the samples under the musicians control were all plain recordings of traditional physical instruments, and could only be played back unprocessed. In v2 longer loops of audio material were given to the musicians, this time using 'pitched' conventional instruments, in this case the guitar. In this version of the instrument the musicians were given some processing control over the raw sample data, being able to shape the playback envelope, control filter resonance and cutoff parameters, and pitch-shift the sample in real-time. Finally, in v3 *more* sample manipulation control was offered to the musicians, including granular synthesis-based time stretching, and also a synthesizer was added to the sample playback options. In taking advantage of the increased control over their sound the musicians moved away from playing back conventional instrumental sounds and towards the synthetic sounds made possible by digital signal processing (see section 7.2.2).

The move beyond natural sounds was an aspect of digital synthesisers which fascinated Deleuze (Gilbert, 2004). As discussed in section 7.1.2, the assemblage is characterised by the new potentials for action it opens up. In the Viscotheque v3 jam sessions, the assemblage of musicians, technology and atmosphere over time moved away from the norms of conventional musicality. Deleuze saw this as the rejection of constraint in sound-making—throwing off the rules about 'what sound could come next'—and that this opening up of new potential sounds was the highest aim to which a creative assemblage could aspire.

One noticeable difference between the v3 jam sessions and the jam session for the previous two versions was the intensity of the atmosphere, particularly in peak moments. The loud, droning synthetic sounds were only achievable by morphing the original samples beyond recognition using the direct audio processing control afforded the musicians through the multi-touch mapping. Indeed, the average loudness of the sessions increased week-to-week as they became more familiar with the interface (see fig. 8.8(c)). The intimate connection to (and identification with) the sound was a theme in the v3 post-jam interviews. This connection and familiarity allowed the musicians to push the sonic envelope, giving rise the these intense affective atmospheres.

The intensity of the task domain in digitally-mediated creativity is perhaps what sets it apart from more prosaic human-computer interaction contexts. Groupware and collaborative computing systems have been around for decades. Often, these systems were designed to facilitate the creation and manipulation of digital artefacts, such as documents, source code and other representations of knowledge. In Viscotheque (and indeed in any digital music environment) the primary artefact is the *sound*—which is ephemeral. Particularly in jamming, where there is no audience and the music is not intended for distribution, the atmosphere and feeling of the jam is all there is. This ephemerality perhaps contributes to the intensity of the sound. Knowing that there is no 'long game' for the sound, only the feeling and affect it engenders here and now, leads to a state of urgency.

In these heightened affective states, the sense of distinction between the musician and their sound collapses, they are *present* in a world of sound. As discussed in section 7.2.1, the sound in these cases seemed to shape the musician's action just as much as they shaped the sound through their instruments. This *felt need* to fit with and move in response to the sound paints a more complicated picture of agency than the cognition-action loop model of the human user in HCI. In closing the 2006 re-release of her milestone *Human-Machine Reconfigurations: Plans and Situated Actions*, Lucy Suchman states:

> The point in the end is not to assign agency either to persons or to things but to identify the materialisation of subjects, objects, and the relations between them as an effect, more and less durable and contestable, of ongoing socio-material practices. (L. Suchman, 2006,

p286)

The results of the Viscotheque jam sessions (particularly v3) bear this out. The musicians were certainly exercising their creativity in jamming with the Viscotheque DMI, but they were also subject to the pull of the sound in a way in which (at least subjectively) they were not in control of their actions but guided by the sonic and affective atmosphere present in the jam session at that time. The direct mapping at least allowed the musicians to respond in-the-moment to these atmospheres, and this is perhaps the primary design lesson to take from the Viscotheque jam sessions.

9.2. Disagreeing about difference

I now return to the issues surrounding experience and evaluation which are such hot topics in contemporary HCI discourse, as discussed in chapter 2. The key point of tension between traditional HCI and the third wave is *the nature and implications of difference*. The holistic view of UX (as described in section 2.3) has at its centre an emphasis on the *felt* experience of the human in interaction. It is a deeply interactionist philosophy in which the relations between past experiences and anticipated futures are fundamental to the outworking of human-computer interaction. Comparisons between *different* users of the same digital artefact are therefore difficult to make, because their experience may be very different depending on their own unique histories and tastes. Humans are not stateless, and each person's personal narrative is different from any other person's, although they may have shared passions and experiences.

In asserting and embracing the uniqueness of any specific configuration of humans and machines in time and space, there is a reluctance to propose normative laws which may hold more generally. The *differences* between even very (outwardly) similar configurations of humans, computers and the environment shape the trajectory of that interaction—these differences cannot be unthinkingly glossed over.

> New techniques in HCI itself are converging to suggest that multiple, potentially competing interpretations can fruitfully co-exist. (Sengers and Gaver, 2006)

The difference in interpretations is to be celebrated; it is not a problem to be dealt with. One rationale for this openness is given by Bødker, one of the leading proponents of the third wave nomenclature:

> many of the questions that we need to deal with as designers of the new multiple, experience-oriented technology are still so open that we need to make technological experiments in order to understand which questions to ask. (Bødker, 2006)

The openness is motivated by the novelty of these new experience oriented technologies, and the fact that we do not yet even know the right questions to ask. The purpose of experimentation is to explore the space of design possibilities, and differences in the response of different users or different environmental variations give crucial insight into the nature of this space. These differences are wild, autonomous, and unplanned, and the surprise of an unexpected variation in user activity or experience is heralded with great joy.

In contrast, in the traditional HCI experimental process, the hypothesis must be known in advance. Then, any differences between experimental conditions are carefully controlled, so as to be able to attribute causal links between those differences and variations in the measured outcomes. Participants in these experiments are selected to be representative samples so that their differences are unbiased (in the statistical sense). Sure, they are invariably different people who will respond in their own subtly unique ways to a given set of circumstances and stimuli, but these variations are expected to be 'normally distributed', with an expectation that these contributions are just noise, and that with a sufficiently large sample size these variations will cancel out and the *true* value of the quantity being measured will emerge. The differences will be smoothed over, and the human-computer interaction problem becomes one of studying the behaviour of the idealised user, the *essence* of the user which is the same across the whole population of individuals.

In this paradigm differences are 'domesticated'; they are under the control of the experimenter. Any differences which are *not* controlled for and specifically manipulated in the different experimental conditions are assumed to not matter, at least as far as the measurable outcomes of the experiment are concerned. This is a sacrifice that the experimenter makes knowingly, not in naïveté to the fact that the users are not all identical, but in the belief that there is a shared substrate which they all share and about which, through the specific logic and reasoning of the experimental process, general laws can be articulated. This is a scientific method akin to the physical sciences, and it allows very specific hypotheses to be evaluated and adjudicated between with a great deal of certainty.

In the third wave approach, however, the differences between individuals, and the influence of other users (in group HCI contexts) and the influence of cultural, environmental and emotional factors are fundamentally *uncontrollable*. What is missing compared to the traditional HCI case is the confidence that the differences are unbiased, that they average out given enough participants. On the contrary, the differences between environments and cultural backgrounds and current mood and hopes and dreams only become more fragmented as more participants are considered. This is perhaps why third wave user studies tend to focus on detailed descriptions of small groups of users—the number of dimensions along which users can differ results in a combinatorial explosion in the amount of descriptive work required, and the overhead of simply differentiating the users becomes prohibitive, let alone trying to uncover any links or patterns between them.

This issue is at the root of the disagreement on measuring UX which divides the holists and the reductionists (as discussed in section 2.3). Both camps agree that the maturation and proliferation of computational devices and environments requires a re-evaluation of the goal of HCI as a discipline. Both agree that experience, rather than usability, is the ultimate goal of interaction and the dimension along which we must optimise. Where they differ is on how to compare *different* experiences, both between different people or between different experiences of the same system across different temporal and spacial scales. The reductionists argue that it is possible to compare different experiences in a given 'equivalence class' (modulo a certain experiment with a specific measurable result), where any differences between the users undergoing that experience are irrelevant to the value of the quantity under inspection. The holists counter that there exist no such equivalence classes, and that any comparison of numerical representations of experience is unable to capture the nuances of the experience which are essential to their very nature. They conclude that the comparison is so impoverished as to be worthless.

It appears that we are at an impasse. HCI traditionalists feel that difference is a tool at their disposal, a weapon to wield in dividing and conquering the rich tapestry of forces and influences which are at work when humans and machines come together. The third wave counters by *problematising* difference—pointing out all the ways in which those differences which the traditionalists purport to be in control of are in fact wild, uncontrollable and free. In the middle of this debate, how does one do research?

HCI practitioners need to be careful about the assumptions about the populations of users they claim to represent in their controlled studies.

> Social and cultural realities are no simply givens; they are performed, enacted, and reproduced in the course of everyday life. (Dourish, 2007)

The differences these social and cultural realities between even participants in the same experiment may be significant, and work must be done to at least be sensitive this in the analysis of results. We must also be wary of the temptation to assign a number to every one of these new dimensions of analysis. If we are at least aware of these factors, we are less likely to sweep these differences under the carpet as we collapse all participants into factors in ANOVA tests. This is not to say that such tests are useless, merely that the act of imposing these categories is an act of representing the world and the differences in it in a certain way.

The third wave, on the other hand, would do well to resist the temptation to uphold novelty and difference as an end in itself. While no two snowflakes are the same, some may be more similar than others. There may be meaningful measures of difference, which will be highly dependent on the particular human-computer interaction task under investigation. Indeed, one of the driving forces behind third wave HCI has been the development of computational environments which support complicated, multi-user, exploratory and creative interaction between users and their environment. This does not mean that all systems are as complex as one another. There is still a great deal of important research to be done in improving usability, particularly in touch interfaces, and in these cases there are meaningful measures of task efficiency. It is difficult to make a judgement call which differences are controllable and which are wild.

9.3. Machine learning in the third wave

Beyond this generic (and obvious) 'both sides can learn from each other' platitude, one way forward which I have attempted to sketch in this thesis (particularly in chapter 8) is the use of Machine Learning (ML)[1] techniques to uncover patterns in the data. Because of the technological progress described in section 2.1, pursuits which fall under the broad umbrella of 'the humanities', and especially the arts, are being infiltrated and augmented with computing devices. Interactive music making, visual and new media art; these are just some of the interaction contexts in which are now HCI's business. And one key implication of this infiltration

[1] See section 8.2 for an overview of the ML process.

is the ease with which we can log and reason about the interaction which happens in these contexts and the differences which emerge between users. It seems shortsighted to reject outright the use of numerical reasoning in favour of the thick descriptions of ethnography.

There are ways to engage in numerical and quantitative analysis which *are* sensitive to the feral nature of difference in human-computer interaction. A richer picture can be obtained by considering *all* the readily available data, and computers are excellent tools for doing so. Indeed, computers are very good good at picking up certain types of differences.[2] ML techniques can be used not just to find things that are the same but to articulate *how they are different*. These differences can then feed into a more holistic approach to understanding the interaction which gave rise to the data.

What I am advocating, then, is the use of ML techniques not for their prediction accuracy but for the insight they provide into the *structure* of the data.[3] Unsupervised techniques such as clustering may be suitable for this purpose, but supervised learning techniques may also be used if the target value (regression) or label (classification) are chosen appropriately. Rather than using subjective labels or measures of experience or artistic quality, use the naturally occurring divisions in the data—in the case of the Viscotheque v3 jam sessions these were the different musicians, different groups and different sessions.

This is why no UX questionnaires were used in the Viscotheque jam sessions. Although such questionnaires are useful tools in certain situations they are an example of the 'measurement and quantification of experience' approach which is so contentious in the UX literature (see section 2.3). Instead, the interaction logs were the only quantitative data considered in the analysis. This led to the discovery of the patterns across the different musicians and groups, particularly the distinctiveness in the way that the musicians held and moved the device. This difference between the musicians opens up many interesting questions for future study, such as the effect of co-location or the use of accelerometer 'signatures' to identify musicians directly from the data.

This approach will not satisfy the most ardent critics of representationalism. The musician (with all their hopes, dreams, influences and aesthetic preferences) must still be represented quantitatively by a feature vector, and the choice of which features to use and which to exclude is far from value-neutral. But it does provide at least a more *objective* approach to quantifying what went on in the jam sessions. This invariably opens up the procedure to accusations of representational error and bias, but at least the criticism of subjectivity on the part of the musician is deflected.

If the burden of representation shifts to the feature vector, it can at least be much richer (from an information-theoretic perspective) than any subjective measure of experience. Any questionnaire longer than (say) one hundred questions will prohibitively time consuming to complete. In comparison, the Viscotheque v3 jam sessions generated 2.7 million OSC packets and this log data weighed in at 297MB.

It would be foolish to say that there is a linear (or indeed any clear mathematical relationship) between the size of the data and the insight it provides. Indeed, getting a feel for and

[2]Babbage's *difference engine* (Swade, 2005) was so named with good reason

[3]Traditional statistics such as (generalised) linear models may also be helpful, although ML is particularly suitable because of its focus on large and multi-modal data.

reasoning about the questionnaire results may be tractable for a human observer, while the sheer amount of data in the logs is overwhelming—but this is precisely the benefit that ML provides. These techniques are capable finding patterns which are *unexpected*. These patterns are then able to be accepted or dismissed based on their compatibility with the bigger picture.

It is true that there are many components of the musicians' interaction (and potential axes of difference) which are much more difficult to instrument. Non-verbal communication, emotional state and aesthetic sensibilities are still as difficult as ever to measure. But it would be foolish to ignore the data we can easily collect, and to not mine it with sophisticated ML techniques to see what differences *do* emerge between musicians. In this, ML finds perhaps an unlikely ally in the affect theory ideas discussed in chapter 7. Affects and affective atmospheres give rise to (or inhibit) *action*, they are concerned with what bodies can *do*. Insofar as the logs capture the movements of the musicians as the interact and the sound and visuals which they are caught up in, they provide a base from which to reason about what was going on. This is also an incentive to measure the musicians more closely, including bio-metrics such as EEG (brain), ECG (heart), EDA (skin) and EMG (skin) (see Nacke et al., 2010).

These are not new insights, but situating them in the context of the disagreements about difference that characterise HCI discourse in this third wave is hopefully helpful in thinking about how to sensitively use mathematical and algorithmic techniques in experientially-oriented HCI contexts. In particular, by collecting as much data as possible and using 'natural' class labels or response variables, it may be possible to better understand *open-ended creative interaction* in post-hoc analysis. The patterns which turn up in this analysis can then guide future analysis, and perhaps even suggest more constrained scenarios which allow stronger causal links to be drawn between the cause and effect of the phenomena of interest. This is not the only way to use ML techniques even in DMI design. Fiebrink et al. (2011) used online (real-time) learning to incorporate musicians into the model training process, and in the age of big data more uses will be found for these techniques.

In light of this grand narrative the actual application of ML in the Viscotheque design process may seem a bit shallow and inconsequential. There is an interesting finding about the accelerometer data, but no clarity as to the cause or implications of this result. What has been lost in the data analysis is specificity and causality. This is largely a result of the fact that the analysis was all post-hoc. The jam sessions were designed to be as open-ended as possible, and all analysis was exploratory in the sense that it was looking for *any* pattern rather than testing a specific hypothesis. It is useful then to see this data-driven approach as a 'hypothesis generation' phase, and that the insights which come out of the open-ended interaction may be testable in more detail with future jam sessions. Whether this approach falls back into the trap of 'domesticating difference' is a difficult question to answer, and any future jam sessions must be designed in a fashion sensitive to this critique.

9.4. Chapter summary

The affective power of sound was a key theme of the Viscotheque post-jam interviews, and this is perhaps one of the key differences between computer-supported collaborative *cre-*

ativity and computer-supported collaborative *work*. The intensity of the atmospheres which characterised the Viscotheque jam sessions indicate a complex relationship between human and machine agency. By moving towards a natural (direct) sound mapping, the Viscotheque DMI attempted to close the loop between the musician and their sound.

In the context of the wider trends in third wave HCI discussed in chapter 2, many of the criticisms of traditional HCI seem to stem from underlying disagreements about which differences are 'controllable' and which are not. This has led some to reject numerical analysis in favour of the 'thick description' of ethnography. However, there are ways to apply numerical ML techniques in a way that is sensitive to the critique of the third wave, and these quantitative techniques can uncover interesting and helpful patterns and relationships which may be missed by human observers. These insights, taken together with expert observations, can hopefully provide a fuller picture of the complex interaction between humans and machines in computer-supported creativity.

10. Conclusion

The Viscotheque DMI was developed with careful consideration of the musical and experiential context of jamming. Over three design iterations the interface evolved from a very simple 'process control' interface in v1 to a more expressive multi-touch sample manipulation tool in v3. At each stage of the design process, open-ended jam sessions held with local musicians suggested that the potential was there for the interface to support rich jamming experiences. The v3 interface, which represents the *current* state of the Viscotheque DMI, was shown to provide able support for free-form improvisation amongst musicians who, despite experience in other jamming contexts, were new to the Viscotheque instrument. This alone is encouraging, and future versions of the instrument will continue to pursue the goal of providing computational support for the practice of jamming.

Aside from the construction of the Viscotheque system, this thesis has attempted to show how DMI design theory and expert judgements can be combined with qualitative and quantitative feedback from field trials in the design of a smartphone-based DMI. By avoiding the measurement of participant experience with subjective quantitative tools such as experience surveys, this design process has hopefully dealt sensitively with the issues surrounding UX which are so contentious in third wave HCI. The affective atmospheres and synthetic timbres in intense jam sessions present a picture of human-computer interaction which is vivid, intense and compelling. These affective atmospheres are increasingly a part of our interaction with (and through) computers as HCI continues to explore the intersection of user experience, the arts and technology. At the same time, the use of ML techniques (particularly in the v3 log analysis) also provided some insights into the interaction data, such as the patterns in the way the different musicians held their devices. This finding was missed in the observation of the musicians on the videotape, and suggests interesting possibilities for future studies.

It may seem as though the themes in this thesis are somewhat schizophrenic, attempting to walk the tightrope between the 'thick descriptions' of ethnography and critical theory and the quantitative results and hard numbers of more traditional HCI. These criticisms have some merit—it *is* a temptation to switch between the narrow focus of quantitative results and the fuzzy, nebulous theory of the third wave as suits our purposes. Still, this type of interdisciplinary research hopefully hints at the way that perspectives from different areas, so often in disagreement in HCI discourse, can be used together in a sensitive way. In particular, I hope that the discussion of 'machine learning in the third wave' (section 9.3) is helpful in this regard.

Music interface research in HCI has long felt the need to justify its design decisions with techniques more suited to technologies in the workplace. Music interaction designers are increasingly liberated to affirm the real reasons we build the tools that we build—the ability of music to bring joy to the heart. We have not forgotten *why* we jam, hopefully we are

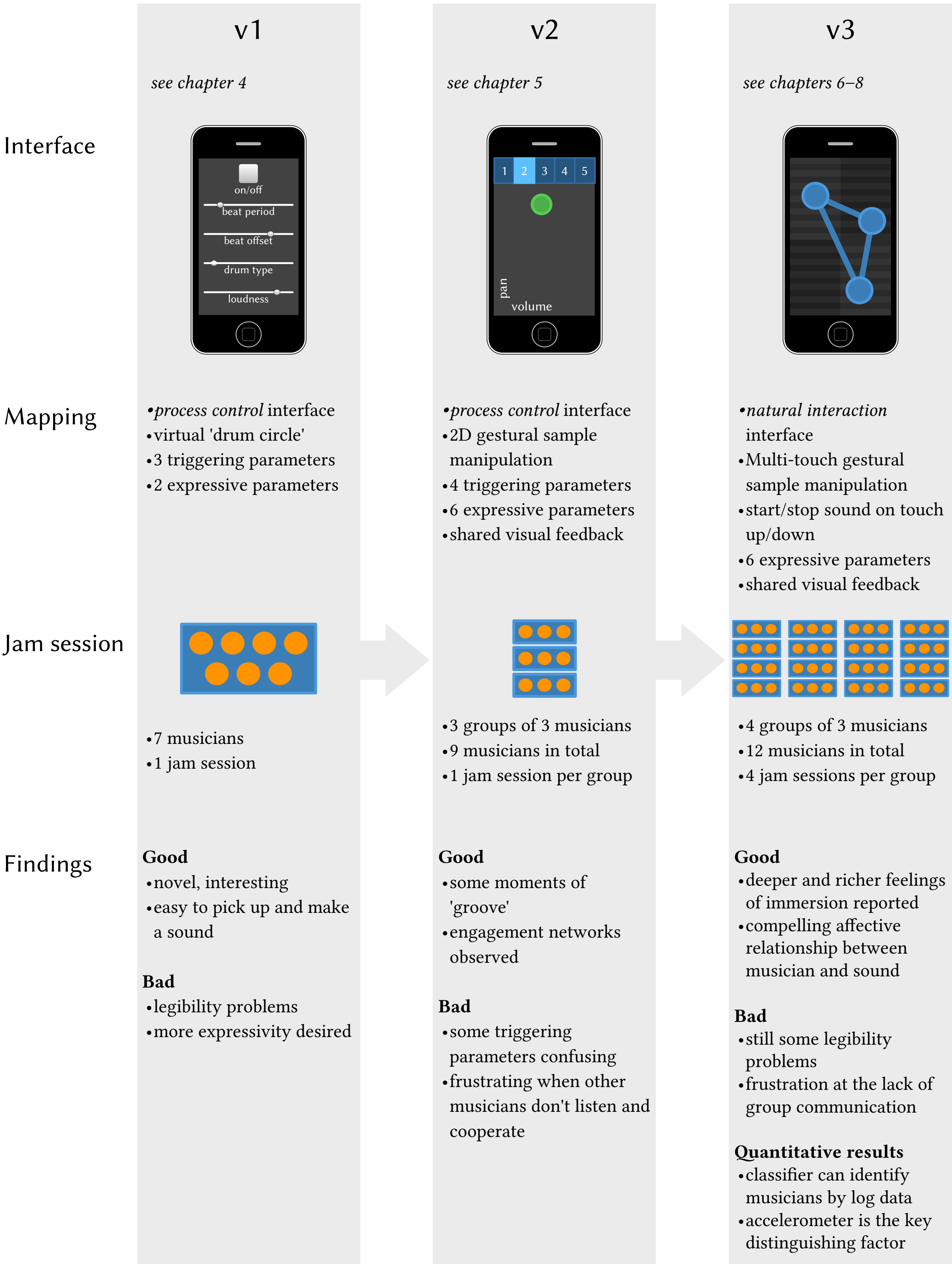


Figure 10.1.: Overview of the Viscotheque design process, including details and findings from the jam sessions held with each version of the instrument.

increasingly able to justify our design decisions in mainstream HCI discourse.

10.1. Summary of contributions

A multi-touch sound mapping for smartphone-based touchscreen DMIs (chapters 4 to 6)
A multi-touch sound mapping was developed and refined over the three versions of the Viscotheque DMI. These refinements were made through a series of field trial jam sessions with trained musicians. These jam sessions were open-ended and exploratory, with the participants able to define their own creative practices in the confines of what was possible with the instrument. Using both expert judgements, feedback from musicians and log data analysis, the improvements to the instrument led to an (at times) very satisfying jamming experience for the musicians in the v3 jam sessions.

A view of affect which affirms the intensity of human-computer interaction in creative contexts (chapter 7)
The term *affect* has a meaning in affect theory (which falls under the broad umbrella of the social sciences) which is different to its usual definition in HCI discourse. By discussing these ideas in the context of human-computer interaction, this thesis promotes a use of the term affect to mean a pre-personal, unstructured *potential* for action. This notion of affect may provide a counterpoint to the 'individualised' picture of behaviour and emotion which third wave HCI has been so critical of, particularly in creative contexts like sound and music computing.

A machine learning approach sensitive to the critique of third wave HCI (chapter 8)
The Viscotheque analysis was performed with the aim of using machine learning techniques on the detailed log data collected in Viscotheque without falling victim to the 'quantitatism' and measurement of experience of which the third wave is so critical. This involved using only partitions of the data which arose naturally, and avoiding any subjective measure of experience quality. Instead of predictive accuracy, the goal of the ML analysis was to expose patterns in the data which were missed in the observation of the jam sessions. The consistency of the accelerometer signatures from each musician was the key finding of this analysis.

A. Viscotheque v3 feature vector

The feature vector is a hybrid feature vector, including features based on the touch, accelerometer and audio data. A listing of the features used is given in section 8.3.

A.1. Mode features

The *instrument mode* features are a representation of the proportion of time the musician spent in each 'mode' of the instrument for that time window. Each mode (which was governed by how many concurrent touches were on the screen) afforded control of a different sonic dimension of the output sound, as discussed in section 6.2. One touch allowed lowpass filtering, two touches for volume and time-stretching, three touches for pitch shifting, and four touches for switching between the sampler and the synthesizer.

Each of these features takes a value in the interval $[0, 1]$, and the values will sum to 1. For example, if for a given 5 second time window the musician was only using the 3 finger pitch shifting mode of the instrument, then the mode 3 feature would have a value of 1 and all the others would have a value of 0.

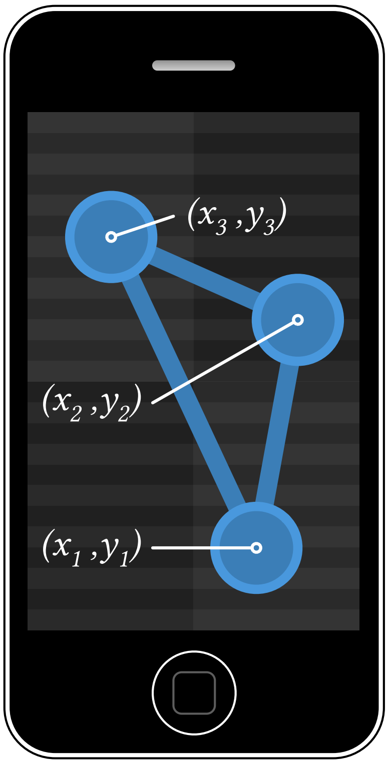
A.2. Touch activity features

These features are based on the touch activity without reference to the position of the touches on the screen. They include the number of 'touch-down's and 'touch-up's per second, the number of touch moves per second, and the proportion of time the musician spent in the different musical modes (i.e. how many fingers were on the screen, see section 6.2 for details). These features were designed to capture the musical aspirations of the musician to the extent that they reflect the *nature* (and vigour) of their musical manipulations.

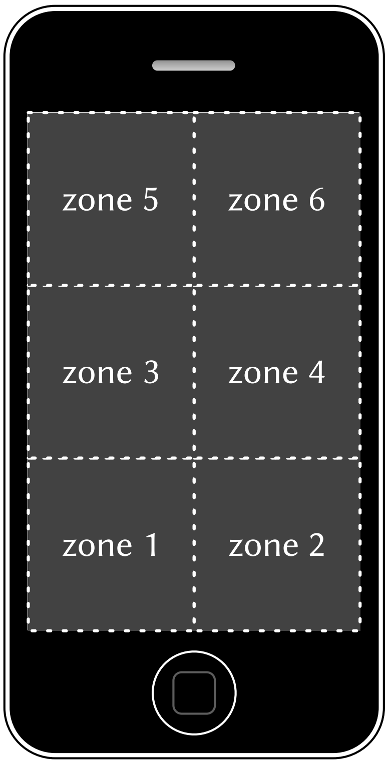
A.3. Touch position features

For each time window, every touch location (both for the initial touch-downs and any touch movements thereafter) has an x and y position as shown in fig. A.1(a). The touch position features measure the mean value and spread of these touches over the time window. For instance, if the total touch region covered spans the iPhone screen from top to bottom, but is narrowly focused on the left hand edge of the screen, then the touch x standard deviation will be low but the touch y standard deviation will be high.

This is a deliberately 'broad' description of the finger touches of the musician—it will capture high-level similarities (such as preferring the top of the screen to the bottom) but does



(a) Each finger touch has a (x, y) co-ordinate in the region $[0, 1] \times [0, 1]$. The touch position features are the mean and s.d. (in both x and y) of all the finger touches in the time window.



(b) Touch zones for the feature vector. The touch position features are a (normalised) histogram representing the time spent in each zone.

Figure A.1.: The touch position and touch zone features.

not discern between subtly different gestures. Partially, this is due to the difficulty of recognising gestures that are not known a priori. Rather than looking for specific gestures, these features are designed (as part of a larger feature vector) to capture general trends in the touch location and variance explored by the musicians during their viscotheque jamming.

A.4. Touch zone features

These features are based on the position of the fingers on the screen. They are designed to complement the touch activity features. Each touch zone feature represents the (normalised) amount of time that zone had at least one finger touch in it. For example, if the musician spent the whole time with a finger in each of the top two corners of the screen (as seems to be the case for group 4, session 4 in fig. 8.1) then the zone 5 and zone 6 features would be 1 and the the other zone features would be 0.

As with the touch position features, the zone breakdown of only six zones is deliberately coarse—there is probably no difference in musical intention between a touch 3mm from the bottom corner and a touch 4mm from the bottom corner, and the features are designed to recognise these touches as the same. Of course, there are points near the touch boundaries where small variations in touch position *do* lead to different representations in the feature vector, but the nature of the touches (primarily smooth arcs rather than individual points) means that this is less of an issue.

Also, the relationship between the finger touch position and the musical intention and expression of the musician at that point is not a straightforward one-to-one map (as discussed in section 6.2). The sound being produced depends on the other fingers on the screen, where the touch *started* on the screen (which may be a different position to where it ends up) and which sample/instrument is being played at the time. The mode features are more meaningful in this regard, and so the touch position features are designed to capture any visual similarity between the musicians on the screen.

A.5. Accelerometer features

These features capture the orientation and movement of the device over the feature window as captured by the device's 3-axis accelerometer. There are seven accelerometer features in total: six of which are the mean value and standard deviation for each of the x, y, and z axes, and a RMS motion energy value. The mean value statistics capture the average position of the device over the time window, while the standard deviation statistics capture how much that position varied. The RMS motion energy statistic is a scalar measure of how much the device's accelerometer vector varied over the time window—when the device is shaken vigorously this value will be large, and when it is stationary this value will be (near) zero.

A.6. Audio features

Finally, we have the audio features used in the feature vector. These audio features were primarily chosen for their perceptual salience, for example the perceptually-scaled loudness value discussed in section 8.1.2. The reasons for this are twofold; firstly because our goal is not classification accuracy for its own sake but a model which exposes the structure in the data in a way that is meaningful, and secondly because the feature window length of 5s, while short compared to human musician timescales involved in the jam, is long compared to the millisecond-length frames common in audio signal processing, and many common audio features become less useful at these long timescales.

The features audio features used are:

- **Octave-band signal intensity ratio (OBSIR)**
  The purpose of the OBSIR is "to capture in a rough manner the power distribution of the different harmonics of a musical sound without recurring to pitch detection techniques" (Essid et al., 2006) This is a timbral feature which is independent of pitch.

- **Zero-crossing rate (ZCR)**
  This feature measures the number of zero-crossings in the raw audio signal. Although this is a very low-level audio feature, it has been shown to be helpful in discriminating between musical and speech-like sounds (Scheirer and Slaney, 1997).

- **Autocorrelation** the total autocorrelation of the audio signal, this is a measure of the periodicity of the signal.

- **Loudness**
  Perceptually scaled aggregate loudness (intensity) of the sound over the time window (Moore et al., 1997).

- **Perceptual spread**
  The spread of the loudness coefficients, a measure of how 'spread out' the low, mid and high frequency components of the sound are (Peeters, 2004).

- **Sharpness**
  Closely related to brightness, sharpness depends on the location of narrow-band peaks in the loudness spectrum (Zwicker and Fastl, 1999).

- **Spectral flatness**
  The degree to which the frequencies present in the signal are uniformly distributed. White noise is maximally flat, while a sinusoid has the minimum possible value. Another feature for discerning between noise-like and musical sounds (Mitrović et al., 2010).

- **Spectral rolloff**
  The frequency that 99% of the spectral energy lies below.

- **Spectral variation**
  The normalised correlation between the spectrum of consecutive frames, this is a measure of how much variation there is in the timbre of the sound over the window.

References

Agre, P. E. (1997). *Computation and Human Experience.* Cambridge University Press (cit. on p. 10).

Ahonen, T. (2011). *Tomi Ahonen Almanac 2011.* Mobile Telecoms Industry Annual Review (cit. on p. 8).

Alterhaug, B. (2004). "Improvisation on a triple theme: Creativity, Jazz Improvisation and Communication". In: *Studia Musicologica Norvegica* 30, pp. 97–118 (cit. on p. 32).

Anderson, B. (2009). "Affective atmospheres". In: *Emotion, Space and Society* 2.2, pp. 77–81 (cit. on p. 101).

Apple Reports First Quarter Results (2012). Apple Inc. URL: http://www.apple.com/pr/library/2012/01/24Apple-Reports-First-Quarter-Results.html (cit. on p. 8).

Arduino. N Zambetti. URL: http://www.arduino.cc/ (cit. on p. 8).

Audio Unit Programming Guide (2007). URL: https://developer.apple.com/library/mac/#documentation/MusicAudio/Conceptual/AudioUnitProgrammingGuide/Introduction/Introduction.html (cit. on p. 44).

Baeg, S.-H., J.-H. Park, J. Koh, K.-W. Park, and M.-H. Baeg (2007). "Building a smart home environment for service robots based on RFID and sensor networks". In: *Control, Automation and Systems, 2007. ICCAS '07. International Conference on*, pp. 1078–1082 (cit. on p. 7).

Ballagas, R., M. Rohs, J. G. Sheridan, and J. Borchers (2008). "The Design Space of Ubiquitous Mobile Input". In: *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, p. 386 (cit. on p. 8).

Bardzell, J. and S. Bardzell (2008). "Interaction Criticism: a Proposal and Framework for a New Discipline of HCI". In: *CHI '08 extended abstracts on Human factors in computing systems.* ACM Request Permissions (cit. on pp. 15, 102).

Bardzell, J. and S. Bardzell (2011). "Pleasure is your birthright: digitally enabled designer sex toys as a case of third-wave HCI". In: *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 5).

Benyon, D., K. Höök, and L. Nigay (2010). "Spaces of interaction". In: *ACM-BCS '10: Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference.* British Computer Society (cit. on p. 7).

Berliner, P. (1994). *Thinking in Jazz : The Infinite Art of Improvisation.* 1st ed. Chicago Studies in Ethnomusicology. University Of Chicago Press (cit. on pp. 25, 34).

Bertelsen, O. W., M. Breinbjerg, and S. Pold (2007). "Instrumentness for creativity mediation, materiality & metonymy". In: *C&C '07: Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition.* ACM Request Permissions (cit. on pp. 5, 15).

Blackman, L. (2008). *The Body (Key Concepts).* First Edition. Berg Publishers (cit. on p. 101).

Blaine, T. and S. Fels (2003a). "Collaborative musical experiences for novices". In: *Journal Of New Music Research* 32.4, pp. 411–428 (cit. on p. 37).

Blaine, T. and S. Fels (2003b). “Contexts of collaborative musical experiences”. In: *NIME ’03: Proceedings of the 2003 Conference on New Interfaces for Musical Expression.* National University of Singapore (cit. on pp. 36, 37, 51, 147).

Blaine, T. and T. Perkis (2000). “The Jam-O-Drum interactive music system: a study in interaction design”. In: *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pp. 165–173 (cit. on pp. 8, 70).

Bloch, C. (2000). “Flow: Beyond Fluidity and Rigidity”. In: *Human Studies* 23.1, pp. 43–61 (cit. on p. 33).

Blythe, M. A. (2004). “Pastiche scenarios”. In: *interactions* 11.5 (cit. on p. 15).

Blythe, M. A., J. J. McCarthy, A. Light, S. Bardzell, P. Wright, J. Bardzell, and A. Blackwell (2010). “Critical dialogue: interaction, experience and cultural theory”. In: *CHI EA ’10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems.* ACM (cit. on p. 15).

Bødker, S. (2006). “When Second Wave HCI meets Third Wave Challenges”. In: *NordiCHI ’06: Proceedings of the 4th Nordic conference on Human-computer interaction.* ACM Press (cit. on pp. 5, 13, 14, 149).

Boehner, K., R. DePaula, P. Dourish, and P. Sengers (2005). “Affect: from information to interaction”. In: *CC ’05: Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*, pp. 59–68 (cit. on pp. 71, 100).

Borgmann, A. (1987). *Technology and the Character of Contemporary Life: A Philosophical Inquiry.* University Of Chicago Press (cit. on p. 38).

Bowers, J. (2002). “Improvising Machines Ethnographically Informed Design for Improvised Electro-Acoustic Music”. PhD thesis. University of East Anglia, Norwich, UK (cit. on p. 29).

Breiman, L. (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32 (cit. on pp. 134, 144).

Breiman, L. (2002). *Manual on setting up, using, and understanding Random Forests v4.0* (cit. on p. 134).

Brennan, T. (2004). “The transmission of affect”. In: *Cornell University* (cit. on p. 101).

Brewer, J., A. Williams, and P. Dourish (2007). “A handle on what’s going on: combining tangible interfaces and ambient displays for collaborative groups”. In: *TEI ’07: Proceedings of the 1st international conference on Tangible and embedded interaction* (cit. on p. 8).

Brodersen, C. and J. F. Kristensen (2004). “Interaction through negotiation”. In: *NordiCHI ’04: Proceedings of the third Nordic conference on Human-computer interaction.* ACM (cit. on p. 11).

Brown, B., S. Reeves, and S. Sherwood (2011). “Into the wild: challenges and opportunities for field trial methods”. In: *CHI ’11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 17).

Bryan-Kinns, N. (2004). “Mobile group music improvisation”. In: *Proc. of Engagability and Design 2004, Birmingham, UK* (cit. on p. 86).

Bryan-Kinns, N. and F. Hamilton (2009). “Identifying Mutual Engagement”. In: *Behaviour & Information Technology* (cit. on p. 85).

Buchanan, I. (2004). “Deleuze and Music”. In: *Deleuze and Music.* Ed. by I. Buchanan and M. Swiboda. Edinburgh University Press (cit. on p. 100).

Burmester, M., M. Mast, K. Jager, and H. Homans (2010). “Valence method for formative evaluation of user experience”. In: *DIS ’10: Proceedings of the 8th ACM Conference on Designing Interactive Systems.* ACM Request Permissions (cit. on p. 15).

Camurri, A., G. Varni, and G. Volpe (2010). “Towards analysis of expressive gesture in groups of users: computational models of expressive social interaction”. In: *Gesture in Embodied Communication and Human-Computer Interaction* 5934, pp. 122–133 (cit. on p. 125).

Cao, X., C. Forlines, and R. Balakrishnan (2007). “Multi-user interaction using handheld projectors”. In: *UIST ’07: Proceedings of the 20th annual ACM symposium on User interface software and technology* (cit. on p. 8).

Caputo, B., K. Sim, and F. Furesjo (2002). “Appearance-based object recognition using SVMs: which kernel should I Use”. In: *NIPS workshop on statistical methods for computational experiments in visual processing and computer vision.* Proc of NIPS workshop on Statistical methods … (cit. on p. 134).

Card, S. K., T. P. Moran, and A. Newell (1986). *The Psychology of Human-Computer Interaction.* New edition. Lawrence Erlbaum Associates (cit. on pp. 5, 10).

Carroll, J. and M. Rosson (2007). “Participatory Design in Community Informatics”. In: *Design Studies* 28, pp. 243–261 (cit. on p. 10).

Church, K. and B. Smyth (2009). “Understanding the intent behind mobile information needs”. In: *IUI ’09: Proceedingsc of the 13th international conference on Intelligent user interfaces* (cit. on p. 8).

Clayton, M., R. Sager, and U. Will (2004). “In time with the music: The concept of entrainment and its significance for ethnomusicology”. In: *ESEM counterpoint* 1 (cit. on p. 29).

Cochrane, T. (2008). “Expression and Extended Cognition”. In: *The Journal of Aesthetics and Art Criticism* (cit. on p. 29).

Colebrook, C. (2003). *Understanding Deleuze.* Australian Cultural Studies. Allen & Unwin (cit. on pp. 33, 104).

Cook, P. (2001). “Principles for designing computer music controllers”. In: *NIME ’01: Proceedings of the 1st Workshop on New Interfaces for Musical Expression.* National University of Singapore, pp. 1–4 (cit. on p. 37).

Costello, B. and E. Edmonds (2009). “Directed and emergent play”. In: *C&C ’09: Proceeding of the seventh ACM conference on Creativity and cognition* (cit. on pp. 13, 14).

Costello, B., L. Muller, S. Amitani, and E. Edmonds (2005). “Understanding the experience of interactive art: Iamascope in Beta space”. In: *IE2005: Proceedings of the second Australasian conference on Interactive entertainment* (cit. on pp. 61, 71).

Create Digital Music. URL: `http://createdigitalmusic.com` (cit. on p. 37).

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience.* Harper Collins New York (cit. on p. 30).

Csikszentmihalyi, M. and K. Rathunde (1993). “The measurement of flow in everyday life: Toward a theory of emergent motivation”. In: *Nebraska symposium on …* (cit. on p. 30).

Cummins, F. (2009). “Rhythm as an Affordance for the Entrainment of Movement”. In: *Phonetica* (cit. on pp. 30, 104).

Dahl, L. and G. Wang (2010). “Sound Bounce: Physical Metaphors in Designing Mobile Music Performance”. In: *NIME ’10: Conference on New Interfaces for Musical Expression* (cit. on p. 37).

Dannenberg, R. B., B. Thom, and D. Watson (1997). “A Machine Learning Approach to Musical Style Recognition”. In: *Proceedings of the International Computer Music Conference*, pp. 344–347 (cit. on p. 123).
DeLanda, M. (2002). *Intensive Science & Virtual Philosophy*. Continuum (cit. on pp. 100, 102).
DeLanda, M. (2006). *New Philosophy of Society: Assemblage Theory and Social Complexity*. Continuum (cit. on p. 103).
Deleuze, G. and F. Guattari (1987). *Thousand Plateaus: Capitalism and Schizophrenia*. 1st ed. University Of Minnesota Press (cit. on p. 102).
Dempsey, N. P. (2008). “Hook-Ups and Train Wrecks: Contextual Parameters and the Coordination of Jazz Interactions”. In: *Symbolic Interaction* 31.1, pp. 57–75 (cit. on p. 25).
Dennis, A. R., R. M. Fuller, and J. S. Valacich (2008). “Media, tasks, and communication processes: A theory of media synchronicity”. In: *MIS quarterly* 32.3, pp. 575–600 (cit. on p. 24).
Dewey, J. (1934). *Art as Experience*. Penguin (cit. on p. 15).
Dewsbury, J. D. (2011). “The Deleuze-Guattarian assemblage: plastic habits”. In: *Area* 43.2, pp. 148–153 (cit. on p. 102).
Dey, A. K., P. Ljungstrand, and A. Schmidt (2001). “Distributed and disappearing user interfaces in ubiquitous computing”. In: *CHI ’01: CHI ’01 extended abstracts on Human factors in computing systems* (cit. on p. 7).
DiSalvo, C., P. Sengers, and H. Brynjarsdóttir (2010). “Mapping the Landscape of Sustainable HCI”. In: *CHI ’10: Proceedings of the 28th international conference on Human factors in computing systems*. ACM Press (cit. on p. 12).
Dobrian, C. and D. Koppelman (2006). “The ’E’ in NIME”. In: *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06)* (cit. on pp. 15, 36).
Dobson, M. (2010). “Performing your self? Autonomy and self-expression in the work of jazz musicians and classical string players”. In: *Music Performance Research*, pp. 42–60 (cit. on p. 35).
Doffman, M. (2009). “Making It Groove! Entrainment, Participation and Discrepancy in the ‘Conversation’of a Jazz Trio”. In: *Language & History* 52.1, pp. 130–147 (cit. on pp. 30, 32).
Dourish, P. (2001a). “Seeking a Foundation for Context-Aware Computing”. In: *Human-Computer Interaction* 16.2, pp. 1–15 (cit. on p. 11).
Dourish, P. (2001b). *Where the Action Is*. The Foundations of Embodied Interaction. The MIT Press (cit. on p. 11).
Dourish, P. (2004). “What we talk about when we talk about context”. In: *Personal and Ubiquitous Computing* 8.1, pp. 19–30 (cit. on p. 12).
Dourish, P. (2007). “Seeing like an interface”. In: *OZCHI ’07: Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces* (cit. on p. 151).
Dubnov, S. (2005). “Improvisation planning and jam session design using concepts of sequence variation and flow experience”. In: *Proc. International Conference on Sound and Music Computing* (cit. on p. 30).
Edmonds, E. and L. Candy (2002). “Creativity, art practice, and knowledge”. In: *Communications of the ACM* 45.10 (cit. on p. 14).

Essid, S., G. Richard, and B. David (2006). “Musical instrument recognition by pairwise classification strategies”. In: *Audio, Speech, and Language Processing, IEEE Transactions on* 14.4, pp. 1401–1412 (cit. on p. 162).

Essl, G. and M. Rohs (2009). “Interactivity for mobile music-making”. In: *Organised Sound* 14.02, pp. 197–207 (cit. on p. 37).

Fallman, D. (2011). “The new good: exploring the potential of philosophy of technology to contribute to human-computer interaction”. In: *CHI ’11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on pp. 5, 6, 11–13).

Fels, S. and M. Lyons (2009). “Creating new interfaces for musical expression: introduction to NIME”. In: *SIGGRAPH ’09: SIGGRAPH 2009 Courses.* ACM (cit. on p. 36).

Fiebrink, R., P. R. Cook, and D. Trueman (2011). “Human model evaluation in interactive supervised learning”. In: *CHI ’11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM Request Permissions (cit. on p. 153).

Finstad, K. (2010). “The Usability Metric for User Experience”. In: *Interacting with Computers* 22.5, pp. 323–327 (cit. on p. 15).

Flores, L., M. Pimenta, E. Miranda, E. Radanovitsck, and D. Keller (2010). “Patterns for the design of musical interaction with everyday mobile devices”. In: *IHC ’10: Proceedings of the 9th Symposium on Human Factors in Computing Systems.* Brazilian Computer Society (cit. on pp. 51, 95).

Forlizzi, J. and K. Battarbee (2004). “Understanding Experience in Interactive Systems”. In: *DIS ’04: Proceedings of the 5th conference on Designing interactive systems*, pp. 261–268 (cit. on p. 15).

Foroughi, H., B. Aski, and H. Pourreza (2008). “Intelligent video surveillance for monitoring fall detection of elderly in home environments”. In: *Computer and Information Technology, 2008. ICCIT 2008*, pp. 219–224 (cit. on p. 7).

Frauenberger, C. and J. Good (2010). “Phenomenology, a framework for participatory design”. In: *Proceedings of the 11th Biennial Participatory Design Conference* (cit. on p. 17).

Furniss, D., A. Blandford, and P. Curzon (2011). “Confessions from a grounded theory PhD: experiences and lessons learnt”. In: *CHI ’11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 11).

Garfinkel, H. (1967). *Studies in Ethnomethodology.* Prentice Hall (cit. on pp. 11, 26).

Gaver, B., T. Dunne, and E. Pacenti (1999). “Design: Cultural probes”. In: *interactions* 6.1 (cit. on p. 18).

Gaye, L., L. Holmquist, F. Behrendt, and A. Tanaka (2006). “Mobile music technology: Report on an emerging community”. In: *NIME* 6, pp. 22–25 (cit. on p. 37).

Genuer, R. and J. Poggi (2010). “Variable selection using random forests”. In: *Pattern Recognition Letters* (cit. on p. 134).

Gilbert, J. (2004). “Becoming-Music: The Rhizomatic Moment of Improvisation”. In: *Deleuze and Music.* Ed. by I. Buchanan and M. Swiboda. Edinburgh University Press (cit. on pp. 103, 109, 148).

Gill, S. (2007). “Entrainment and musicality in the human system interface”. In: *AI & Society* (cit. on p. 29).

Gilmore, D. J., G. Cockton, E. Churchill, S. Kujala, A. Henderson, and M. Hammontree (2008). "Values, Value and Worth: Their Relationship to HCI?" In: *CHI EA '08: CHI '08 extended abstracts on Human factors in computing systems* (cit. on p. 12).

Gostner, R., E. Rukzio, and H. Gellersen (2008). "Usage of spatial information for selection of co-located devices". In: *MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (cit. on p. 8).

Greenhough, B. (2011). "Assembling an island laboratory". In: *Area* 43.2, pp. 134–138 (cit. on p. 103).

Gregg, M. and G. J. Seigworth, eds. (2010). *The Affect Theory Reader.* Duke University Press Books (cit. on p. 100).

Guest, G., K. MacQueen, and E. Namey (2011). *Applied Thematic Analysis.* Sage Publications, Inc (cit. on p. 99).

Gurevich, M., P. Stapleton, and A. Marquez-Borbon (2010). "Style and Constraint in Electronic Musical Instruments". In: *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)* (cit. on p. 37).

Halpern, M., J. Tholander, M. Evjen, S. Davis, A. Ehrlich, K. Schustak, E. Baumer, and G. Gay (2011). "MoBoogie: creative expression through whole body musical interaction". In: *CHI '11: Proceedings of the 2011 Conference on Human Factors in Computing Systems.* ACM (cit. on p. 37).

Hamilton, A. (2007). "Music and the Aural Arts". In: *The British Journal of Aesthetics* (cit. on p. 22).

Haraway, D. (1991). "A Cyborg Manifesto". In: *Simians, Cyborgs and Women: The Reinvention of Nature.* Routledge, New York, pp. 149–181 (cit. on p. 109).

Hardt, M. (1993). *Gilles Deleuze: An Apprenticeship in Philosophy.* 1st ed. University Of Minnesota Press (cit. on p. 101).

Harrison, S., D. Tatar, and P. Sengers (2007). "The three paradigms of HCI". In: *Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems San Jose, California, USA* (cit. on pp. 5, 19).

Hassenzahl, M. (2008). "User experience (UX): towards an experiential perspective on product quality". In: *IHM '08: Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine.* ACM Request Permissions (cit. on p. 15).

Hassenzahl, M., S. Diefenbach, and A. Göritz (2010). "Needs, affect, and interactive products - Facets of user experience". In: *Interacting with Computers* 22.5, pp. 353–362 (cit. on pp. 15, 19).

Hassenzahl, M. and N. Tractinsky (2005). "User experience—a research agenda". In: *Behaviour & Information Technology* 25.2, pp. 91–97 (cit. on p. 14).

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics).* 2nd ed. 2009. Corr. 3rd printing 5th Printing. Springer (cit. on p. 134).

Hinrichs, U. and S. Carpendale (2011). "Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits". In: *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 17).

Hintze, J. (1998). "Violin Plots: A Box Plot Density Trace Synergism". In: *American Statistician* 52 (cit. on p. 56).

Hu, W., T. Dinh, P. Corke, and s. Jha (2010). "Design and Deployment of Long-term Outdoor Sensornets: Experiences from a Sugar Farm". In: *Pervasive Computing, IEEE* 99, p. 1 (cit. on p. 7).

Irani, L., J. Vertesi, P. Dourish, K. Philip, and R. E. Grinter (2010). "Postcolonial computing: a lens on design and development". In: *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems.* ACM Request Permissions (cit. on p. 12).

Ishii, H. and B. Ullmer (1997). "Tangible bits: towards seamless interfaces between people, bits and atoms". In: *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM Press, pp. 234–241 (cit. on p. 8).

Jaimes, A. and N. Sebe (2007). "Multimodal human–computer interaction: A survey". In: *Computer Vision and Image Understanding* (cit. on p. 8).

Jain, R. (2008). "Multimedia information retrieval: watershed events". In: *MIR '08: Proceeding of the 1st ACM international conference on Multimedia information retrieval* (cit. on p. 119).

Jennett, C., A. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton (2008). "Measuring and defining the experience of immersion in games". In: *International Journal of Human-Computer Studies* 66, pp. 641–661 (cit. on p. 14).

Jordà, S., G. Geiger, M. Alonso, and M. Kaltenbrunner (2007). "The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces". In: *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction* (cit. on p. 8).

Kaptelinin, V., B. Nardi, S. Bødker, J. Carroll, J. Hollan, E. Hutchins, and T. Winograd (2003). "Post-cognitivist HCI: second-wave theories". In: *CHI EA '03: CHI '03 extended abstracts on Human factors in computing systems* (cit. on p. 6).

Kiefer, C. and N. Collins (2008). "HCI methodology for evaluating musical controllers: A case study". In: *NIME '08: New Interfaces for Musical Expression* (cit. on p. 5).

Kjeldskov, J. and C. Graham (2003). "A review of mobile HCI research methods". In: *Human-computer interaction with mobile devices and services* 2975, pp. 317–335 (cit. on pp. 7, 8).

Knight, P. (2004). "Towards the Limit-Experience in Music Performance". PhD thesis. University of Melbourne (cit. on p. 32).

Knörig, A. (2007). "The bodily aspect in computer-supported creativity". In: *C&C '07: Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition.* ACM Request Permissions (cit. on p. 15).

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions.* University of Chicago Press (cit. on p. 5).

Lamont, A. (2009). "Strong experiences of music in university students". In: *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)*, pp. 250–259 (cit. on p. 32).

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-Network-Theory.* Oxford University Press (cit. on p. 103).

Law, E. L.-C. (2011). "The measurability and predictability of user experience". In: *EICS '11: Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems.* ACM Request Permissions (cit. on pp. 15–17).

Law, E. L.-C., V. Roto, M. Hassenzahl, A. Vermeeren, and J. Kort (2009). “Understanding, scoping and defining user experience: a survey approach”. In: *CHI ’09: Proceedings of the 27th international conference on Human factors in computing systems* (cit. on p. 15).

Lawrence, T. (2006). “In Defence Of Disco (Again)”. In: *New Formations* 58.Summer, pp. 128–146 (cit. on p. 102).

Leong, T. W., P. Wright, F. Vetere, and S. Howard (2010). “Understanding experience using dialogical methods: the case of serendipity”. In: *OZCHI ’10: Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction.* ACM Request Permissions (cit. on p. 16).

Lewis, P. (2006). *Book Review: Thinking in Jazz.* University of California Los Angeles (cit. on p. 34).

Liaw, A. (2002). “Classification and Regression by randomForest”. In: *R News* 2.3, pp. 18–22 (cit. on p. 134).

Light, A. (2011). “Digital interdependence and how to design for it”. In: *interactions* (cit. on p. 17).

Light, A. (2006). “Adding method to meaning: A technique for exploring peoples’ experience with technology”. In: *Behaviour & Information Technology* 25.2, pp. 175–187 (cit. on p. 98).

Linden, J. van der, R. Johnson, J. Bird, Y. Rogers, and E. Schoonderwaldt (2011). “Buzzing to play: lessons learned from an in the wild study of real-time vibrotactile feedback”. In: *CHI ’11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 17).

MacDonald, R. (2005). “Musical identities of professional jazz musicians: a focus group investigation”. In: *Psychology of Music* 33.4, pp. 395–417 (cit. on pp. 23, 35).

MacDonald, R., C. Byrne, and L. Carlton (2006). “Creativity and flow in musical composition: an empirical investigation”. In: *Psychology of Music* (cit. on p. 30).

MacDonald, R. and B. Wilson (2006). “Constructions of jazz: How jazz musicians present their collaborative musical practice”. In: *Musicae Scientiae* 10.1, pp. 59–83 (cit. on pp. 23, 27, 31).

Malbon, B. (1999). *Clubbing: Dancing, Ecstasy, Vitality.* Critical Geographies. Routledge (cit. on p. 102).

Mamykina, L., L. Candy, and E. Edmonds (2002). “Collaborative creativity”. In: *Communications of the ACM* 45.10 (cit. on p. 14).

Mankoff, J. C., E. Blevis, A. Borning, B. Friedman, S. R. Fussell, J. Hasbrouck, A. Woodruff, and P. Sengers (2007). “Environmental sustainability and interaction”. In: *CHI ’07: CHI ’07 extended abstracts on Human factors in computing systems* (cit. on p. 12).

Maslow, A. H. (1954). *Motivation and Personality.* Harper & Brothers (cit. on p. 15).

Massumi, B. (1995). “The autonomy of affect”. In: *Cultural Critique* (cit. on p. 100).

Massumi, B. (2002). *Parables for the Virtual: Movement, Affect, Sensation.* Duke University Press Books (cit. on p. 101).

Mathews, M. (1963). “The Digital Computer as a Musical Instrument”. In: *Science* 142.3592, pp. 553–557 (cit. on p. 35).

Mazzola, G. (2008). “Jazz and the Creative Flow of Collabortive Gesturtes”. In: *Annäherungen und Grenzüberschreitungen* (cit. on p. 32).

McCarthy, J. J. and P. Wright (2004). *Technology as Experience.* MIT Press (cit. on pp. 15, 16).

McDermott, J., D. Sherry, and U. O'Reilly (2011). "Creative Transformations: How Generative and Evolutionary Music can Inform Music HCI". In: *When Words Fail: What can Music Interaction tell us about HCI? Workshop at BCS HCI 2011* (cit. on p. 37).

Merritt, T., W. Kow, C. Ng, K. McGee, and L. Wyse (2010). "Who makes what sound?: supporting real-time musical improvisations of electroacoustic ensembles". In: *OZCHI '10: Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction.* ACM Request Permissions (cit. on p. 43).

Meyer, D. (2001). "Support Vector Machines". In: *R News* 1.3 (cit. on p. 133).

Miranda, E. and M. Wanderley (2006). *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard.* 1st. Computer Music and Digital Audio Series. A-R Editions, Inc. (cit. on pp. 3, 36).

Mitrović, D., M. Zeppelzauer, and C. Breiteneder (2010). "Features for Content-Based Audio Retrieval". In: *Advances in Computers* 78, pp. 71–150 (cit. on p. 162).

Monk, A., M. Hassenzahl, M. A. Blythe, and D. Reed (2002). "Funology: designing enjoyment". In: *CHI '02: CHI '02 extended abstracts on Human factors in computing systems* (cit. on p. 14).

Monson, I. (1997). *Saying Something: Jazz Improvisation and Interaction.* 1st ed. University Of Chicago Press (cit. on pp. 23, 25, 35, 36).

Montuori, A. (2003). "The complexity of improvisation and the improvisation of complexity: Social science, art and creativity". In: *Human Relations* 56.2, p. 237 (cit. on p. 34).

Moore, B. C. J., B. R. Glasberg, and T. Baer (1997). "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness". In: *J. Audio Eng. Soc* 45.4, pp. 224–240 (cit. on pp. 123, 162).

Moore, G. E. (1975). "Progress in digital integrated electronics". In: *Electron Devices Meeting* 21, pp. 11–13 (cit. on p. 7).

Morrison, A., S. Viller, and P. Mitchell (2011). "Building sensitising terms to understand freeplay in open-ended interactive art environments". In: *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 14).

Nacke, L., M. Grimshaw, and C. Lindley (2010). "More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game". In: *Interacting with Computers* 22.5, pp. 336–343 (cit. on p. 153).

Nardi, B. A. (1995). *Context and Consciousness: Activity Theory and Human-Computer Interaction.* The MIT Press (cit. on p. 11).

Nesbitt, N. (2010). "Critique and Clinique: From Sounding Bodies to the Musical Event". In: *Sounding the Virtual: Gilles Deleuze and the Theory and Philosophy of Music.* Ed. by B. Hulse and N. Nesbitt. Ashgate Publishing Company, p. 304 (cit. on pp. 38, 104, 109).

New Oxford English Dictionary (2010). 3rd ed. Oxford University Press (cit. on p. 21).

Nielsen, J. (1993). *Usability Engineering.* Morgan Kaufmann (cit. on p. 14).

Norgaard, M. (2011). "Descriptions of Improvisational Thinking by Artist-Level Jazz Musicians". In: *Journal of Research in Music Education* 59.2, pp. 109–127 (cit. on p. 28).

Norman, D. A. (1988). *The Psychology Of Everyday Things.* Basic Books (cit. on p. 5).

Norman, D. A. (2005). *Emotional Design: Why We Love (or Hate) Everyday Things.* Basic Books (cit. on p. 14).

O'Brien, H. and E. Toms (2008). "What is user engagement? A conceptual framework for defining user engagement with technology". In: *Journal of the American Society for Information Science and Technology* 59.6, pp. 938–955 (cit. on p. 16).

Oh, J., J. Herrera, N. Bryan, L. Dahl, and G. Wang (2010). "Evolving The Mobile Phone Orchestra". In: *NIME '10: Proceedings of the 2010 Conference on New Interfaces for Musical Expression* (cit. on p. 37).

O'Modhrain, S. (2011). "A Framework for the Evaluation of Digital Musical Instruments". In: *Computer Music Journal* 35.1, pp. 28–42 (cit. on pp. 15, 36).

Paine, G. (2010). "Towards a Taxonomy of Realtime Interfaces for Electronic Music Performance". In: *NIME '10: Proceedings of the 2010 Conference on New Interfaces for Musical Expression* (cit. on p. 36).

Patel, N., J. Clawson, A. Voida, and K. Lyons (2009). "Mobiphos: A study of user engagement with a mobile collocated-synchronous photo sharing application". In: *International Journal of Human-Computer Studies* 67.12, pp. 1048–1059 (cit. on p. 85).

Peeters, G. (2004). *A large set of audio features for sound description (similarity and description) in the cuidado project.* IRCAM (cit. on p. 162).

Peters, C., G. Castellano, and S. de Freitas (2009). "An exploration of user engagement in HCI". In: *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, pp. 1–3 (cit. on p. 82).

Picard, R. and G. Cosier (1997). "Affective intelligence—the missing link?" In: *BT Technology Journal* 14.4 (cit. on p. 100).

Pinheiro, R. (2011). "The creative process in the context of jazz jam sessions". In: *Journal of Music and Dance* 1.1, pp. 1–5 (cit. on pp. 24, 26, 34).

Poppe, R. and R. Rienks (2007). "Evaluating the future of HCI: challenges for the evaluation of emerging applications". In: *Artifical Intelligence for Human Computing*, pp. 234–250 (cit. on p. 7).

Postel, J. (1980). *RFC 768: User Datagram Protocol.* Internet Engineering Task Force. URL: `http://www.ietf.org/rfc/rfc768` (cit. on p. 44).

Potter, J. and A. Hepburn (2005). "Qualitative interviews in psychology: problems and possibilities". In: *Qualitative research in Psychology* 2.4, pp. 281–307 (cit. on p. 78).

Pousman, Z., M. Romero, A. Smith, and M. Mateas (2008). "Living with tableau machine: a longitudinal investigation of a curious domestic intelligence". In: *UbiComp '08: Proceedings of the 10th international conference on Ubiquitous computing*. ACM Request Permissions (cit. on p. 18).

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing.* ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria (cit. on p. 113).

Redlinger, E. *MRMR - Open Mobile Touch Protocol.* URL: `http://mrmr.noisepages.com` (cit. on pp. 46, 61).

Robson, D. (2001). "PLAY!: sound toys for the non musical". In: *NIME '01: Proceedings of the 1st Workshop on New Interfaces for Musical Expression.* National University of Singapore, pp. 1–3 (cit. on p. 37).

Rode, J. A. (2011). "Reflexivity in digital anthropology". In: *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 5).

Rogers, Y. (2004). “New Theoretical Approaches for Human-Computer Interaction”. In: *ARIST: Annual Review of Information Science and Technology*, pp. 87–142 (cit. on pp. 8, 9, 18).

Russ, S. W. (1993). *Affect and Creativity: the Role of Affect and Play in the Creative Process.* Personality Assessment Series. Psychology Press (cit. on p. 100).

Satchell, C. (2006). “Cultural Theory: From Armchair Critic to Star Performer”. In: *OZCHI ’06: Proceedings of the 18th Australian Conference on Computer-Human Interaction.* ACM (cit. on p. 100).

Satchell, C. (2008). “Cultural theory and real world design: Dystopian and Utopian Outcomes”. In: *CHI ’08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems.* ACM Request Permissions (cit. on pp. 15, 100).

Sauro, J. and J. R. Lewis (2009). “Correlations among prototypical usability metrics: evidence for the construct of usability”. In: *CHI ’09: Proceedings of the 27th international conference on Human factors in computing systems.* ACM Request Permissions (cit. on p. 14).

Sawyer, R. K. (2003). *Group Creativity: Music, Theater, Collaboration.* Psychology Press (cit. on p. 22).

Sawyer, R. K. (2004). “Improvised Lessons: Collaborative discussion in the constructivist classroom”. In: pp. 1–14 (cit. on p. 24).

Sawyer, R. K. (2006). “Group creativity: Musical performance and collaboration”. In: *Psychology of Music* 34.2, p. 148 (cit. on p. 30).

Sawyer, R. K. (2007). *Group Genius: The Creative Power of Collaboration.* Basic Books (cit. on p. 25).

Sawyer, R. K. and S. DeZutter (2009). “Distributed creativity: How collective creations emerge from collaboration”. In: *Psychology of Aesthetics, Creativity, and the Arts* 3.2, pp. 81–92 (cit. on p. 25).

Schaik, P. van and J. Ling (2008). “Modelling user experience with web sites: Usability, hedonic value, beauty and goodness”. In: *Interacting with Computers* 20.3, pp. 419–432 (cit. on p. 15).

Scheirer, E. and M. Slaney (1997). “Construction and evaluation of a robust multifeature speech/music discriminator”. In: *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, pp. 1331–1334 (cit. on p. 162).

Schuler, D. and A. Namioka, eds. (1993). *Participatory Design: Principles and Practices.* CRC / Lawrence Erlbaum Associates (cit. on p. 12).

Sengers, P., K. Boehner, M. Mateas, and G. Gay (2008). “The disenchantment of affect”. In: *Personal and Ubiquitous Computing* 12.5 (cit. on p. 16).

Sengers, P. and B. Gaver (2006). “Staying Open to Interpretation:Engaging Multiple Meanings in Design and Evaluation”. In: *DIS ’06: Proceedings of the 6th conference on Designing Interactive systems.* ACM Press, pp. 99–108 (cit. on p. 149).

Sheldon, K. M., A. J. Elliot, Y. Kim, and T. Kasser (2001). “What is satisfying about satisfying events? Testing 10 candidate psychological needs.” In: *Journal of Personality and Social Psychology* 80.2, pp. 325–339 (cit. on p. 15).

Sheller, M. and J. Urry (2006). “The new mobilities paradigm”. In: *Environment and Planning A* (cit. on p. 8).

Sheridan, J. G., N. Bryan-Kinns, and A. Bayliss (2007). “Encouraging witting participation and performance in digital live art”. In: *BCS-HCI ’07: Proceedings of the 21st British CHI Group*

Annual Conference on HCI 2007: People and Computers XXI: HCI...but not as we know it 1 (cit. on p. 15).

Shouse, E. (2005). “Feeling, Emotion, Affect”. In: *M/C Journal* 8.6 (cit. on p. 100).

Smith, D. and J. Protevi (2008). “Gilles Deleuze”. In: *Stanford Encyclopedia of Philosophy*. Ed. by E. Zalta (cit. on pp. 100, 103).

Sorensen, A. and H. J. Gardner (2010). “Programming with time: cyber-physical programming with impromptu”. In: *OOPSLA ’10: Proceedings of the ACM international conference on Object oriented programming systems languages and applications* (cit. on p. 44).

SoundCloud. URL: `http://soundcloud.com` (cit. on p. 37).

Spinoza, B. (1677). *Ethics*. Penguin Classics. Penguin Classics (cit. on p. 101).

Springett, M. (2009). “Evaluating Cause and effect in User Experience”. In: *Digital Creativity* 20.3, pp. 197–204 (cit. on p. 19).

Steinberg, D. and S. Cheshire (2005). *Zero Configuration Networking: The Definitive Guide*. 1st ed. O’Reilly Media (cit. on p. 64).

Strauss, A. C. and J. M. Corbin (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. 2nd. Sage Publications, Inc (cit. on p. 11).

Suchman, L. (2006). *Human-Machine Reconfigurations: Plans and Situated Actions*. 2nd ed. Cambridge University Press (cit. on pp. 10, 11, 148).

Suchman, L. A. (1987). *Plans and Situated Actions*. The Problem of Human-Machine Communication. Cambridge University Press (cit. on p. 10).

Su, N. M. (2010). “Street Fighter IV: Braggadocio Off and On-line”. In: *CSCW ’10* (cit. on p. 14).

Sussman, G. and G. Steele (1975). “Scheme: an Interpreter for the Extended Lambda Calculus”. In: *MIT AI Lab Memo* (cit. on p. 44).

Suykens, J. A. K. (2001). “IMTC 2001. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No.01CH 37188)”. In: *IMTC 2001. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics*. IEEE, pp. 287–294 (cit. on p. 134).

Swade, D. D. (2005). “The Construction of Charles Babbage’s Difference Engine No. 2”. In: *IEEE Annals of the History of Computing* 27.3, pp. 70–78 (cit. on p. 152).

Swallow, D., M. A. Blythe, and P. Wright (2011). “Grounding Experience: Relating Theory and Method to Evaluate the User Experience of Smartphones”. In: *CHI ’11: Proceedings of the 2011 annual conference on Human factors in computing systems*. ACM (cit. on p. 16).

Swift, B. (2011). “Impish Grooves”. In: *Computer Music Journal* 35.4, pp. 119–137 (cit. on p. iii).

Swift, B. (2012a). “Becoming-Sound: Affect and Assemblage in Improvisational Digital Music-Making”. In: *CHI ’12: Proceedings of the International Conference on Human Factors in Computing Systems*. (Cit. on pp. iii, 99).

Swift, B. (2012b). “Chasing a Feeling: Experience in Computer Supported Jamming”. In: *Music and Human-Computer Interaction*. Springer (cit. on pp. iii, 23).

Swift, B., H. J. Gardner, and A. Riddell (2009). “Distributed Performance in Live Coding”. In: *ACMC ’09: Proceedings of the Australasian Computer Music Conference* (cit. on pp. iii, 41).

Swift, B., H. J. Gardner, and A. Riddell (2010). “Engagement Networks in Social Music-making”. In: *OZCHI ’10: Proceedings of the Australasian Conference on Computer-Human Interaction* (cit. on pp. iii, 61).

Swift, B., H. J. Gardner, and A. Riddell (2011). "A Chasing After the Wind: Experience in Computer-Supported Group Music-Making". In: *When Words Fail: What can music interaction tell us about HCI? Workshop at BCS HCI 2011* (cit. on p. iii).

Tanaka, A. (2002). "Multimodal interaction in music using the Electromyogram and relative position sensing". In: *NIME '02: Proceedings of the 2002 Conference on New Interfaces for Musical Expression* (cit. on p. 37).

Tanaka, A. (2006). "Interaction, experience and the future of music". In: *Consuming Music Together* (cit. on p. 37).

Tanaka, A. (2010). "Mapping Out Instruments, Affordances, and Mobiles". In: *Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010)* (cit. on p. 38).

Taylor, A. (2011). "Out there". In: *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM Request Permissions (cit. on pp. 5, 12, 13).

Taylor, R., G. Schofield, J. Shearer, J. Wallace, P. Wright, P. Boulanger, and P. Olivier (2011). "Designing from within: humanaquarium". In: *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on pp. 16, 58).

Telesco, P. and D. Waterhawk (2003). *Sacred Beat: From the Heart of the Drum Circle.* Red Wheel Weiser (cit. on p. 41).

Tomkins, S. (1963). *Affect Imagery Consciousness Volume I: The Positive Affects.* Springer (cit. on p. 101).

Tuedio, J. (2006). ""And Then Flew On": Improvisational Moments of Rhizomatic Assemblage in Grateful Dead Musical Experience". In: *Texas/Southwest Popular Culture Association/American Popular Culture Association Joint National Conference* (cit. on p. 27).

Twidale, M., X. Wang, and D. Hinn (2005). "CSC: computer supported collaborative work, learning, and play". In: *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pp. 687–696 (cit. on p. 15).

Urry, J. (2004). "Small worlds and the new'social physics'". In: *Global Networks* (cit. on p. 24).

Väänänen-Vainio-Mattila, K. (2009). "Developing an expert evaluation method for user eXperience of cross-platform web services". In: *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era* (cit. on p. 15).

Varni, G. (2008). "Emotional entrainment in music performance". In: *Automatic Face & Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, pp. 1–5 (cit. on p. 29).

Verplank, B. and C. Sapp (2001). "A course on controllers". In: *NIME '01: Proceedings of the 1st Workshop on New Interfaces for Musical Expression* (cit. on p. 37).

Vihavainen, S., S. Mate, L. Seppälä, F. Cricri, and I. D. D. Curcio (2011). "We want more: human-computer collaboration in mobile social video remixing of music concerts". In: *CHI '11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 8).

Wanderley, M. and N. Orio (2002). "Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI". In: *Computer Music Journal* 26.3, pp. 62–76 (cit. on p. 37).

Weiser, M. (1991). "The Computer for the Twenty-First Century". In: *Scientific American* 265.3, pp. 94–104 (cit. on p. 7).

Wickham, H. (2009). *ggplot2.* Elegant Graphics for Data Analysis. Springer (cit. on p. 113).

Wilkie, K., S. Holland, and P. Mulholland (2010). "What can the language of musicians tell us about music interaction design?" In: *Computer Music Journal* 34.4, pp. 34–48 (cit. on p. 95).

Williams, A., E. Kabisch, and P. Dourish (2005). “From interaction to participation: configuring space through embodied interaction”. In: *UbiComp’05: Proceedings of the 7th international conference on Ubiquitous Computing*. Springer-Verlag (cit. on p. 12).

Winograd, T. and F. Flores (1986). *Understanding Computers and Cognition: A New Foundation for Design.* Ablex (cit. on pp. 10–12).

Wolf, T. V., J. A. Rode, J. Sussman, and W. A. Kellogg (2006). “Dispelling “design” as the black art of CHI”. In: *CHI ’06: Proceedings of the SIGCHI conference on Human Factors in computing systems.* ACM Request Permissions (cit. on pp. 5, 16).

Wolpaw, J., N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan (2002). “Brain–computer interfaces for communication and control”. In: *Clinical Neurophysiology* 113.6, pp. 767–791 (cit. on p. 8).

Wright, M. (2005). “Open Sound Control: an enabling technology for musical networking”. In: *Organised Sound* 10.03, p. 193 (cit. on pp. 41, 91).

Wright, P., J. Wallace, and J. J. McCarthy (2008). “Aesthetics and experience-centered design”. In: *Transactions on Computer-Human Interaction (TOCHI* 15.4 (cit. on p. 16).

Ylirisku, S., V. Halttunen, J. Nuojua, and A. Juustila (2009). “Framing design in the third paradigm”. In: *CHI ’09: Proceedings of the 27th international conference on Human factors in computing systems* (cit. on p. 5).

Young, J. and C. Matheson (2000). “The Metaphysics of Jazz”. In: *Journal of Aesthetics and Art Criticism* 58.2, pp. 125–133 (cit. on p. 24).

Zeileis, A., K. Hornik, A. Smola, and A. Karatzoglou (2004). “kernlab - An S4 Package for Kernel Methods in R”. In: *Journal of Statistical Software* 11.9 (cit. on p. 134).

Zhou, Y., G. Percival, X. Wang, Y. Wang, and S. Zhao (2011). “MOGCLASS: evaluation of a collaborative system of mobile devices for classroom music education of young children”. In: *CHI ’11: Proceedings of the 2011 annual conference on Human factors in computing systems.* ACM (cit. on p. 37).

Zwicker, E. and H. Fastl (1999). *Psychoacoustics: Facts and Models.* 2nd updated ed. Springer Series in Information Sciences. Springer (cit. on p. 162).