Exploring User-Driven Techniques for the Design of New Musical Interfaces through the Responsive Environment for Distributed Performance

Dalia El-Shimy

Department of Electrical and Computer Engineering
McGill University
Montreal, Canada

November 2014

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Abstract

The problem of enriching distributed communication has been tackled extensively by conventional investigations of videoconferencing technology. However, standard tools for computer-mediated communication have yet to offer a level of social engagement that compares to the gold standard of face-to-face interaction. While a number of systems attempt to bridge this gap by providing reasonable support for non-verbal cues, such systems typically attempt to improve collaboration of a practical nature. As such, we were interested in exploring the extent to which computer-mediated technology could support human activities necessitating a greater level of creativity, playfulness and spontaneity. An additional motivation was understanding how distributed collaboration could improve on its co-present counterpart, by leveraging its underlying technology towards further assisting target users in effectively accomplishing the activity at hand.

As musical performance can be considered creative activity necessitating multiple levels of communication and interaction among its participants, we decided to investigate the topics above within the context of distributed musical performance. Given that Computer-Supported Cooperative Work research has illustrated that successful collaboration over a network is contingent not only on resolving technological complexities, but also on the development of appropriate interaction paradigms, we opted to undertake our research from a user- rather than technology-driven perspective. The project was initially propelled by a traditional user-centered design methodology. As such, we began by developing a thorough understanding of our target end user, the musician, by means of contextual observations, interviews, and persona profiles. The information we acquired subsequently inspired a number of simple prototype designs that, in turn, were used in formal user experiments to validate the basic premise of augmenting distributed performance. Eventually, we increased the level of user involvement through a long-term deployment and collaboration with a three-piece ensemble, and a participatory design cycle with a composer.

The final result of our design and development efforts is a responsive environment that augments distributed performance with dynamic, real-time, hands-free control

over several aspects of the musicians' sound, enabling them to seamlessly change volume, affect reverb, adjust the mix, and perceive spatialized audio rendering, without detaching themselves from their higher-level activity. Furthermore, a derivative of this system that provides such features within the context of musical composition was also developed.

Our user-driven design of a novel interactive musical system was not without its share of difficulties. The non-utilitarian nature of the users' tasks poses special challenges, requiring attention to benchmarks, evaluation techniques and alternatives to formal quantitative testing that are suitable to the exacting nature of musical performance. Such challenges are by no means unique to the context of musical performance, but inherent to many creative and artistic domains. As such, this dissertation contributes two novel artefacts—a responsive environment for distributed performance, and a responsive environment for composition—along with a set of recommendations based on our experiences working with a unique and creative end user whose needs cannot easily be defined. In turn, our solutions may be of help to developers looking to acquire a deeper understanding of the user experience that the traditional notion of usability alone does not afford.

Résumé

Le problème de l'enrichissement de la communication distribuée a été largement abordé par les enquêtes classiques de la technologie de vidéoconférence. Cependant, les outils standards pour la communication médiatisée par ordinateur n'offrent pas encore un niveau d'engagement social qui compare a l'étalon-or de l'interaction face-à-face. Bien qu'un certain nombre de systèmes tentent de combler cet écart en fournissant un soutien raisonnable pour les indices non-verbaux, ces systèmes tentent généralement d'améliorer la collaboration de nature pratique. Pour cette raison, nous nous sommes intéressés à explorer la mesure dans laquelle la technologie médiatisée par ordinateur puisse soutenir les activités humaines qui nécessitent un plus grand niveau de créativité, ludisme et de spontanéité. Une motivation supplémentaire était notre désir de comprendre la manière dont laquelle la collaboration distribuée puisse surpasser son homologue du cas co-présent, en profitant de sa technologie sous-jacente pour aider les utilisateurs à accomplir leurs activités plus effectivement.

Comme la performance musicale peut être considérée une activité créative qui nécessite plusieurs niveaux de communication et d'interaction entre ses participants, nous avons décidé d'enquêter les sujets ci-dessus dans le cadre de la performance musicale distribuée. Puisque la recherche sur le travail coopératif assisté par ordinateur a démonstré que le succès de la collaboration sur un réseau depend non seulement sur la résolution des complexités de la technologie, mais aussi sur le développement des paradigmes d'interactions appropriés, nous avons opté d'entreprendre notre recherche selon une perspective axée sur l'utilisateur, plutôt que sur la technologie. Le project était initialement propulsé par la méthodologie de conception centrée sur l'utilisateur traditionelle. Nous avons ensuite haussé le niveau de participation des utilisateurs à travers un déploiment et collaboration à long terme avec un ensemble de trois pièces, et un cycle de conception participative avec un compositeur.

Le résultat final de nos efforts de conception et de développement est un système qui augmente la performance musicale distribuée avec des contrôles dynamique, aux mains libres, en temps réel sur plusieurs aspects du son musical. Notre système permet aux musiciens de changer leur volumes, d'ajuster leurs niveaux de réverbération,

de régler leur mixage, et de percevoir des effets audio spatialisés, tout d'une façon transparente qui ne nécessite pas qu'ils se détachent de leur activité de niveau supérieure. En outre, un dérivé de ce système qui fournit ces mêmes caractéristiques dans le contexte de la composition musicale a également été développé.

Notre conception axée sur l'utilisateur d'un système musical interactif n'était pas sans part sa part de défis. Le caractère non-utilitaire des tâches des utilitsateurs pose des difficultés particulières, exigeant une attention spéciale aux indices de référence et techniques d'évaluation, et des alternatives aux tests quantitatifs formels qui sont adaptés à la nature de la performance musicale. Ces défis ne sont pas uniques au context musical, mais sont aussi inhérent à plusieurs domaines créatifs et artistiques. En conséquence, cette thèse contribue deux objects fabriqués nouveaux—un environnement réactif for la performance distribuée et un environnement réactif for la composition musicale—accompaniés par une série de recommendations basées sur nos éxperiences en travaillant avec des utilisateurs créatifs et uniques, les besoins de qui ne peuvent pas être facilement définis. À leur tour, nos solutions pourrait être utiles aux développeurs qui cherchent à acquérir une compréhension plus profonde de l'expérience de l'utilisateur que la notion traditionelle d'utilisabilité ne peut pas offrir.

Dedication

To Hala and Reda

Acknowledgments

Big up to all my homies who (somehow) managed to put up with me through this insane process, even though they probably wanted to kill me at times. You know how you are.

I'd also like to thank my supervisor Jeremy R. Cooperstock, and the members of my committee, Frank Ferrie and Marcelo M. Wanderley, for their invaluable guidance and advice; Thomas Hermann for inspiring me to get the ball rolling; Steve Cowan for being a wonderful collaborator; the members of People on Pause and the Visitation for stepping up to the plate when I needed to run one too many of those last-minute sessions; and the current and past members of the Shared Reality Lab, who offered immense support while tolerating my non-stop ranting for the past six years.

Contents

List of Figures

List of Tables

List of Acronyms

API	Application Programming Interface
CAASAP	Control Augmented Adaptive System for Audience Participation
CCRMA	Center for Computer Research in Music and Acoustics
CDMI	Collaborative Digital Musical Interaction
CSCW	Computer-Supported Collaborative Work
DIP	Distributed Immersive Performance
DIYSE	Do It Yourself Smart Experience
DML	Digital Music Link
EDAL	Ensemble Delay Acceptance Limit
EDholak	Electronic Dholak
EPM	
External Preference Mapping	
EPT	Ensemble Performance Threshold
FMOL	Faust Music On-Line
GEQ	Gaming Experience Questionnaire
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HSV	Hue, Saturation, and Value
II	Interactive Installation
IR	Infrared
ISM	Interactional Sound and Music
ISX	Internet Sound Exchange

IMN	Interconnected Musical Network
LAA	Latency-Accepting Approach
MIT	Massachusetts Institute of Technology
MCG	Master Cue Generator
NIME	New Interfaces for Musical Expression
Ninjam	Novel Intervallic Jamming Architecture
NMI	New Musical Interface
NMP	Network Musical Performance
OpenCV	Open Source Computer Vision Library
OSC	Open Sound Control
PBSR	Personal Beat Shift Range
PD	Participatory Design
PSE	Problem Solving Environment
PrEmo	Product Emotion Measure
QoE	Quality of Experience
RGT	Repertory Grid Technique
RPI	Rensselaer Polytechnic Institute
RJA	Realistic Jam Approach
RRA	Remote Recording Approach
SARC	Sonic Arts Research Centre
SEI	Sensual Evaluation Instrument
Shira	Structured Hierarchical Interviewing for Requirement Analysis
SoundWIRE	Sound Waves Over the Internet From Real-Time Echoes
SVTE	Shared Virtual Table Environment
TCP	Transmission Control Protocol
TUIO	Tangible User Interface Object
TOUCHE	Task-Oriented and User-Centred Process Model for Developing Inter
UCD	User-Centered Design
UDP	User Datagram Protocol
ViMiC	Virtual Microphone Control
VPS	Videoconferencing Privacy System

VST	Virtual Studio Technology

Chapter 1

Introduction

According to Ackerman, one of the challenges central to the field of Computer-Supported Cooperative Work (CSCW) can be described as the "social-technical gap", a mismatch resulting from the flexible and nuanced nature of human activity when contrasted with the rigid and brittle nature of technical systems [1]. Thus, the author continues, bridging this gap through computational entities (e.g. information transfer, roles, and policies) that are also flexible and nuanced in nature, is essential to the successful design of CSCW applications. This is particularly crucial for distributed collaborative environments, where participants often suffer from a lowered sense of shared awareness, and a decrease in mutual perception of non-verbal cues (e.g., gaze direction, gestures, posture) [171]. Such a problem has been tackled extensively by conventional investigations of videoconferencing technologies: telepresence systems, shared virtual table environments (SVTE) and mobile remote presence (MRP) systems have all emerged in a bid to enrich social engagement within the distributed context. However, such systems strive to improve collaborations of a functional nature, helping to improve cooperation on specific, work-related tasks among remote participants. In an effort to explore the breadth of human activity that computer-mediated communication can enrich, we became particularly interested in examining the creative, ludic and spontaneous aspects of social interaction within a distributed context. An additional motivation was the question of whether distributed collaboration could improve on its co-present counterpart by leveraging its underlying technology towards further assisting target users in effectively accomplishing the activity at hand. One area particularly suited for such investigations, given its socially and temporally exacting nature, is that of distributed musical performance, more formally referred to as network musical performance (NMP).

As CSCW research illustrates, however, successful collaboration over a network is contingent not only on resolving technological challenges, but also on the development of interaction paradigms that can support both the complexities and subtleties of cooperative behaviour. As such, many researchers approach the design of distributed collaborative environments from the user- rather than technology-driven perspective commonly advocated in human-computer interaction (HCI) research. As a temporally exacting activity, demanding multiple levels of communication between the players [47], we argue that the design of novel interactive musical tools can benefit from, and should be afforded, the same level of attention to user needs and behaviours. The only caveat, however, is that user-centered methodologies require careful consideration when applied within the context of musical performance, for reasons discussed below.

Experts in music technology research have for long acknowledged the benefits their field stands to gain from HCI research. Tanaka, for instance, argues that "[i]nstrumental music...establishes rich forms of human-machine interaction that catalyze human-human interaction". Thus, the author continues, the successful design of musical interfaces should be "the result of a fusion of computer-human interface design and acoustic instrument lutherie" [179]. User-centered techniques in particular have enjoyed a growing popularity among developers of new musical interfaces (NMIs)—a term by which we describe novel interactive music systems, gestural controllers, sound installations and sonic environments—keen on improving and sustaining performer and audience engagement. However, the design of NMIs poses some rather interesting problems for traditional HCI techniques. First, the performer is a very unique type of user: his "needs" can be difficult to establish, given that novel artistic tools typically do not exist to a serve concrete purpose, the way a document editor does for example. His "goals" when using such tools can also be too ambiguous to define, seeing as he perhaps has never considered alternatives to his traditional gear. Such challenges are in fact encountered by many interface designers looking to augment creative activities via computing technology: from visual to performance arts, or even the more creative facets of graphic design or photo editing, developing a thorough understanding of the artist-machine relationship is no trivial task. Furthermore, as with any physically and mentally demanding activity, the nature of musical performance imposes strict constraints on any interaction design: a musician's hands are almost always busy playing an instrument. Naturally, his auditory channel is occupied, listening to the sounds he and the other musicians produce. His visual channel is less burdened, but still serves an important role in communicating with his peers. In addition, unless he is a laptop music performer, it is highly unlikely that he will detach himself from the performance and step over to a mouse and keyboard. As a result, many considerations of usability design bear an added level of complexity, and many traditional input and output paradigms become unsuitable.

We decided to examine such challenges by taking a user-driven approach to the design of a novel environment for NMP. By choosing an application area where communication is strongly driven by creativity, self-expression and spontaneity, we hoped to explore the ways in which CSCW systems could better support the "highly flexible, nuanced, and contextualized" aspects of human activity [1]. In addition, we hoped that lessons extrapolated from our efforts could be of use to other developers interested in working on non-utilitarian systems or in creative domains. Given that distributed performance, like many on-line activities, exhibits a decreased sense of sociability among participants [106], we wanted the primary goal of our system to be that of restoring the social aspects of performance. In addition, as Corness and Schiphorst explain, "[p]erformers tacitly know how to pay close attention to bodily cues that accompany movement, as they have consciously developed their awareness of these cues to enabled skilled interaction with other performers" [55]. Thus, we hoped that capitalizing on embodied performer-performer interactions would offer the added advantage of enabling musicians to use our system's functionality without detaching themselves from the higher level task of performance. Furthermore, we wanted to encourage musicians to delve into new sonic territories. By creating a system that allows them to experiment with paradigms that traditional performance does not support, we hoped to make the concept of distributed performance more alluring. Determining the type of functionality our system should afford, however, was no trivial task. We had a number of criteria in mind. First, as a means of extending the social aspects of traditional ensemble music into NMP, we wanted the system to be driven by the interpersonal interactions between distributed musicians. Second, we wanted to offer performers unprecedented control over aspects of their instrumental mix at any given time. Finally, we wanted all controls to be easy to learn and use, and all mappings to be transparent, offering a clear link between user input and system output.

Our efforts resulted in an augmented distributed performance environment that allows musicians to utilize common gestures and behaviours, such as head tilting, body turning and simple motion, as a means of affecting each other's volume and reverb levels, adjusting audio mixes and experiencing spatialized sound. The system was designed for relaxed performance settings that include room for improvisation or experimentation (e.g., loose rehearsals or jams). An example use case scenario for our system would involve geographically displaced friends who wish to play music together over a network, but seek alternatives to traditional videoconferencing that can further enrich their interpersonal interactions. Our performance environment can currently only support electric or electronic (rather than acoustic) instruments, in order to ensure that the modified audio mix played back through the musicians' headphones is not overshadowed by the actual sound of their instruments.

Our user-centric design of a novel, interactive performance environment was met with several challenges. First, given that the vast majority of musicians have never partaken in distributed performance, even a carefully developed understanding of the target users could not help us anticipate the types of interaction they would find useful or the problems they may encounter. Furthermore, the standard notion of usability is not particularly suited to our context, given that the benchmarks of musical performance tend to be hedonic, subjective qualities such as enjoyment, creativity and self-expression. Finally, we found the insight gained from formal, multi-user, quantitative experiments to be rather limited, as they lacked the depth of knowledge necessary to guide us towards significantly improving the overall user experience with our system. These factors encouraged us to develop user-centric approaches, benchmarks and evaluation criteria tailored to the unique nature of musical performance, as described throughout the remainder of this dissertation. As stated earlier, NMI designers have long benefited from HCI knowledge. Nonetheless, we believe that the lessons drawn from taking a purely user-centric approach to the design of novel musical controllers and environments can also be of great use to interface developers wishing to approach creative applications in a rigorous fashion.

Chapter 2

Background and Previous Works

2.1 Computer-Supported Cooperative Work

In 1984, Grief and Cashman invited researchers from varied disciplines to explore the ways in which newly emerging office automation techniques could better support workplace practices [86]. It was during this workshop that the term "Computer-Supported Cooperative Work" was first coined, paving the way for a new "design oriented research area" that has since continued to grow [42]. While Cartensen and Schmidt contend that some debate may surround the exact definition of CSCW, the authors broadly describe the field as one aiming to address "how collaborative activities and their coordination can be supported by means of computer systems" [42]. The vast majority of CSCW systems, also often referred to as groupware, continue to be categorized according to the matrix introduced by Johansen in 1988, and seen here in Figure 2.1 [102]. In essence, collaborations may take place between remote and/or co-located participants, in a synchronous or asynchronous fashion.

Early and popular examples of CSCW include the media spaces developed at the Xerox Palo Alto Research Center (PARC) in the mid-1980s. Described as "[a]n electronic setting in which groups of people can work together, even when they are not resident in the same place at the same time", media spaces allowed participants to create real-time visual and acoustic environments spanning several, physically remote areas. The images and sounds produced in such environments could also be recorded

Time and space-based views of CSCW technologies.

	One meeting site (same places)	Multiple meeting sites (different place)
Synchronous communications (same time)	**Face to Face Interactions** • *Public computer displays* • *Electronic meeting rooms* • *Group decision support systems*	**Remote Interactions** • *Shared view desktop conferencing systems* • *Desktop conferencing with collaborative editors* • *Video conferencing* • *Media spaces*
Asynchronous communications (different times)	**Ongoing Tasks** • *Team rooms* • *Group displays* • *Shift work groupware* • *Project management*	**Communication and Coordination** • *Vanilla e-mail* • *Asynchronous conferencing, bulletin boards* • *Structured messaging systems* • *Workflow management* • *Version control* • *Meeting schedulers* • *Cooperative hypertext, organizational memory*

Fig. 2.1: Computer-Supported Cooperative Work matrix, originally proposed by Johansen, and later updated by Baecker et al., adapted from reference [9]

for later access [20]. As part of the Ontario Telepresence Project, Buxton et al. later merged the idea behind media spaces with the tenets of ubiquitous computing. The result, dubbed Ubiquitous Video, or UbiVid, helped expand interactions between remote participants from dedicated areas to entire workspaces [32]. For instance, Buxton re-configured his office to include various locations where remote visitors could appear: on his desk for one-on-one work, at the coffee table where he holds informal meetings or above his door to check his availability. One of the visiting researchers on the Ontario Telepresence Project was Ishii, who also made significant contributions to early CSCW research. In particular, Ishii introduced the notion of "shared workspaces" as systems that "overcome space and time constraints, and support dynamic collaboration in a work group" [99]. Shared workspaces are designed as a continuous extension of individual work areas, allowing for a seamless, two-way transition between collaborative and individual modes of work. Ishii first exemplified this notion through the TeamWorkStation, a system that not only fused images of its users' desktops with their computer screens, but also allowed them to share their favourite tools from either spaces with other collaborators. The concept of shared workspaces was also a foundation of Ishii's ClearBoard, a drawing medium that allowed remote participants to collaborate over a drawing board using colour markers, electronic pens and natural gestures, all while maintaining direct eye contact with one another [101].

While such early works allowed their developers to explore the various facets of an emerging field, it became particularly apparent to them that the success of CSCW systems is contingent on more than simply novel technology. The developers of media spaces, for instance, explain that "[t]echnologies to support collaborative work are defined by the social setting and by the nature of the work, as well as by the features of the technology" [20]. Such an observation is supported by one of the key findings of the Ontario Telepresence Project, whose chief sociologist stressed that "[t]echnology fails often for social, not technical reasons" [31]. Finally, Ishii et al. highlighted their choice of a user rather than technology-driven design approach, in a bid to create systems that "respect the skills that people use in everyday life" [100]. Such views were in fact distilled by Bannon, who found early CSCW systems to be particularly instrumental in uncovering certain "fictions" about how people integrate tools into their work, and instead encouraged researchers to focus their attention on the nature of collaboration within various settings and contexts [11]. As such, successful CSCW systems must evolve from a thorough understanding of the ways people work and collaborate with one another. Ackerman, for instance, states that "CSCW systems need to have at their core a fundamental understanding of how people really work and live in groups, organizations, communities, and other forms of collective life" [1]. Such a view is supported by Rodden and Blair, who emphasize that "CSCW researchers must focus their efforts to understand and account for the characteristics of cooperative work", as well as Cartensen and Schmidt, who argue that CSCW necessitates a "much better and well conceptualized understanding of cooperative work and its complexity" [42, 161]. Similarly, Sauppé and Mutlu argue that the design of CSCW systems that can effectively support remote collaboration must be informed by "a better understanding of how people collaborate face-to-face and the mechanisms that they use to coordinate their actions" [165].

It therefore comes as no surprise that a number of researchers advocate the adoption of user-driven techniques when it comes to designing CSCW applications, with Bannon and Schmidt going as far as warning that a technology-driven approach could "dilute" the field [13]. Gross et al. add that since CSCW consists, in fact, of a collection “socio-technical systems”, focusing solely on the development of technical features “ignores the potential influence of distributed information of users. . . and their behavior” [84]. As a result, a number of researchers have attempted to improve the design of CSCW systems through in-depth studies of user behaviours and gestures during collaborative activities. For instance, Penichet et al. proposed a Task-Oriented and User-Centred Process Model for Developing Interfaces for Human-Computer-Human Environments, or TOUCHE, in response to the shortcomings of traditional software engineering analysis models within the context of CSCW. The authors argue that while such models may uncover “static and behavioural issues”, they are inadequate when it comes to modelling collaboration. As a result, TOUCHE offers an iterative approach that places the participant-within-a-group at the center of the development process [148]. As another example, Sauppé and Multu modelled the predictive relationship between facial, gestural and vocal cues under dyadic interaction in an effort to understand how such cues affect perceived task success during collaborative activities such as cooperation, instruction and negotiation. In turn, the authors hoped such a comprehensive understanding of social cues might “inform the design of collaboration tools that provide support for a wider range of social cues and that adapt to the changing priorities of communication across different tasks” [165]. Similarly, Cornelius et al. developed a simple framework that characterizes existing CSCW tools according to the mechanisms they use to transmit virtual gestures, the roles of the participants within collaboration and the specific task domain. In addition, the authors used this framework as a basis for a study examining whether surrogate gestures, which may be conveyed by drawing circles and arrows around objects on a shared screen, or natural gestures, which utilize real-time videos of a distant collaborator’s hands or body, are more suited to distributed collaborative design tasks. Perhaps unsurprisingly, their study found that the latter performed significantly better in reducing the distributed users’ cognitive load [54]. Such results are also consistent with the views of Kirk et al., whose study demonstrated that the support of natural remote gestures is critical in helping distributed participants develop a “common ground” during object-focused collaborative tasks. The authors further argue that such an understanding of how remote gestures influence the grounding process can have significant implications to the design of CSCW tools [114].

Interestingly, a number of authors view participatory design, also known as cooperative design and later described in detail in Section 2.4, as simply an extension of CSCW [42]. Kyng, for instance, describes cooperative design as an "instance of cooperative work", making computer support for cooperative design, by extension, an "instance of computer support for cooperative work" [118]. Others, however, such as Bannon, describe such a view as a "mistake that can only add to confusion surrounding both fields" [11]. In fact, Bannon further argues that while user involvement is important to the development of successful CSCW systems, the use of participatory design techniques "does not automatically signify any focus on cooperative work". Thus, it is important to make a distinction between participatory or cooperative design, and the design of systems for cooperative work. Instead, as Cartensen proposes, participatory design should be viewed as a separate tradition that should be applied to the design of CSCW applications if possible [42].

Designing CSCW applications from a user-driven perspective, however, is not without difficulties. The success of an application designed with a focus on its intended users is typically established by means of system evaluations conducted with said users. However, while it may be relatively easy to evaluate the perceptual, cognitive or motor variables that may be central to a single-user application in a laboratory setting, the creation of an ecologically valid scenario for the evaluation of a collaborative system can be far more challenging [85]. Furthermore, Penichet et al. argue that the development of CSCW applications not only requires an understanding of the interactions of the users with a system, but also of their interactions among themselves [148]. This is made particularly difficult by the fact that group work typically evolves intermittently over longer stretches of time, as dictated by its participants' schedules and availabilities, putting the accuracy of observations made during a single, fixed time slot into question. As such, Grudin explains that "[t]ask analysis, design, and evaluation are never easy, but they are considerably more difficult for CSCW applications than for single-user applications" [85].

One last concern in CSCW that we would like to touch upon, given its relevance to our work, is that of awareness. Described by Dourish and Bly as the ability to know "who is 'around', what activities are occurring, who is talking with whom", awareness allows for the informal and spontaneous interactions that are key to maintaining and improving working relationships, but that are often diminished within the distributed context [61]. However, Cartensen describes mutual awareness as "[t]he obvious and fundamental way to coordinate, align, mesh, etc. myriad interdependent and yet distributed activities" [42]. As a result, Mills argues that CSCW researchers should seek to improve our ability to achieve such awareness among participants working through computers and across networks [137]. Awareness entails a certain level of transparency among remote collaborators, allowing them to develop a sense of trust and community that, in turn, encourages the playful and creative sides of interaction. Furthermore, Gross et al. argue that awareness helps increase the orientation of individuals within a group, thereby allowing them to make use of contextual information to accomplish tasks more effectively [84]. As such, it is an aspect of co-operative work that is particularly relevant to our interest in exploring distributed performance, a context where musicians lose the mutual physical perception that is central to successful musical collaboration.

The importance of awareness during performance was also tackled by Fencott and Bryan-Kinns, who notably examined the application of various CSCW principles to co-located musical collaboration [70]. Proposing the term Collaborative Digital Musical Interaction (CDMI) to describe "the phenomenon of technologically supported group musical interaction", the authors sought to explore questions of ownership, territory and privacy, along with the roles that various participants may play while collaborating, and the level of awareness required to support their activities [71]. Furthermore, Bryan-Kinns et al. introduced the notion of Interactional Sound and Music (ISM) to denoting "multiple people interacting together using audio as the primary modality" and, in comparing systems that can support such interaction to standard CSCW applications, the authors also expressed the importance of mutual awareness for successful collaboration [28]. Barbosa has also explored the applicability of CSCW research to the musical domain, in a bid to better understand and support the collaborative aspects of distributed performance that are often hampered by network delays. In fact, the author offers a classification system for "Computer-Supported Cooperative Music" that is heavily inspired by Rodden's CSCW matrix, and even argues that developing paradigms specific to network musical collaboration can lead to novel sonic systems that "express interesting new artistic results" [5].

Finally, we note that a common application area for the "same time/different place" CSCW category that is directly relevant to our research interests is that of videoconferencing, a topic we further explore in the following section.

2.1.1 From Videoconferencing to Telepresence

Videoconferencing systems, or groupware that aims to connect geographically displaced participants through audio and video transmissions, are among some of the most common CSCW applications. While early systems entailed expensive proprietary software and hardware, the introduction of Internet Protocol (IP) based videoconferencing in the 1990s provided access to the public at a relatively low cost. Tools such as Skype or iChat further popularized desktop videoconferencing, allowing face-to-face connections for millions of users around the world to become an everyday reality. Nonetheless, many argue that current videoconferencing systems, and particularly commercial ones, have failed on their promise to support meaningful distributed social engagement. While some attribute the problem in part to the limitations in video and/or audio quality that may arise from insufficient bandwidth [46, 108], we argue that a far more significant problem is the inability of such systems to preserve the "rich set of social behaviours and cues that we as humans know and share" [2]. Eisert, for instance, describes standard videoconferencing as "limited in its support of natural human-centered communication", before adding that the support they offer for cues such as body postures, subtle movements, gaze direction, room acoustics, joint interactions, eye contact and other forms of non-verbal communication tends to be problematic, lacking or entirely absent [64]. To this, Sirkin adds that "[w]e use embodied non-verbal communications such as gestures, body movements, posture, visual orientation, and spatial behavior in concert with our verbal communication to signal our attention, express emotions, convey attitudes, and encourage turn-taking, and... we (perhaps subconsciously) prefer that our technological counterparts follow suit" [171].

The Hydra system, developed by Buxton et al. under the Ontario Telepresence Project described in the previous section, attempted to address issues such as gaze and spatial awareness. The system used independent communication devices dubbed Hydra units, each with its own video display, microphone and speaker, distributed on a local participant's desk to allow for the spatial and acoustical separation of remote collaborators [170]. In essence, such a configuration affords participants the same spatial relationship they would benefit from if they were physically co-present around a table. As such, the Hydra systems allows them to take advantage of "many of the spatial cues of gaze awareness, head turning, gaze awareness and turn taking that are found in face-to-face meetings" [32]. A number of subsequent telepresence systems, or tools that aim to confer a higher level of co-presence than standard videoconferencing systems, took their inspiration from the Hydra system. In particular, shared virtual table environments (SVTE) emerged to give participants the impression of being seated together around a table, thereby allowing them, as Kauff et al. explains, to "make use of rich communication modalities as similar as possible to those used in a face-to-face meeting (e.g., gestures, eye contact, gaze awareness, realistic images, correct sound direction, etc) and eliminate the limits of non-immersive teleconferencing, which impoverish communication (e.g., face-only images in separate windows, unrealistic avatars, no eye contact)" [109]. Such systems also have the added advantage of allowing for greater physical context than desktop videoconferencing systems: participants are aware of their relative placements and, as such, have a mutual physical reference frame [78]. Examples of such systems include TELEPORT, a teleconferencing system that merges a real local environment with a virtual one to give its participants the illusion of sharing the same space [78], and the Virtual Team User Environment (VIRTUE), a "tele-cubicle" that provides a seamless transition between a real desk and a virtual conference table [109].

Such systems are particularly notable for expanding on the notion of shared workspaces proposed by Ishii, and described in the previous section. While Ishii envisioned workers sharing designated real and electronic work areas, SVTEs allow them to share entire rooms. By creating a seamless transition between the real and the virtual, shared virtual table environments give users the impression of being part of an extended perception space, and allow remote participants to be rendered under a correct perspective view in the virtual environment [109], two factors which contribute to an increased sense of co-presence. As an example, Gibbs et al. describe the contribution of the TELEPORT system as its ability to mimic, using 3D modelling and rendering, a shared physical context, and to provide life-sized display of remote participants placed within a virtual space [78]. As such, displaced participants may benefit from the same awareness of consensus regarding objects surrounding them and the relative distance between them, which, in turn, confers a greater level of immersion in the shared space.

A more recent trend in telepresence solutions aiming to transfer the richness of face-to-face interactions to the distributed context is that of telerobotics, also referred to as telepresence robots, mobile remote presence (MRP) systems or embodied proxies. In essence, such systems combine “a live video representation of the remote worker with a local physical platform, often with human body-like proportions” [171]. Like the Hydra system or SVTEs, telepresence robots aim to restore the common social cues and behaviours that typically allow us to perceive and communicate such feelings as engagement, trust and persuasion [2]. Examples of such systems include MeBot, a telerobot that not only communicates audio and video, but also expressive gestures, body pose and proxemics [2]; the Texai Alpha system, which consists of a mobile base, touchscreen, microphone, speakers, pan-tilt webcam, wide-angle camera, and two laser range finders [123]; and Sirkin and Ju’s embodied proxies, which connected a hemispherical base to a video screen by means of an articulating “neck” that could pan and lift at the hemisphere, or tilt the screen itself, thereby mimicking common head motions [171]. Another example is TeleHuman, which, although not mobile, was designed to support greater awareness of gaze and pointing gestures. TeleHuman comprises a life-sized cylindrical display, onto which 360° 3D video models of remote users, captured through multiple Microsoft Kinect units, can be rendered with perspective correction and stereoscopy using an off-the-shelf 3D projector. By preserving 360° motion parallax as a viewer moves around the cylinder, TeleHuman is able to support gaze awareness and gestural interaction beteween distributed participants [113].

By supporting non-verbal cues and gestures, the distributed systems described above are able to approximate face-to-face communication to a greater extent than desktop videoconferencing systems, and, as such, are suitable for a wide variety of collaborative tasks that extend beyond the context of the workplace. The TELEPORT system, for instance, was used by musicians in Geneva to rehearse with a conductor in Germany as part of the Distributed Video Production project [78]. In fact, many systems for network musical performance take their inspiration from videoconferencing tools. Examples of such distributed performance systems, among several others, are thoroughly detailed in the next section.

2.2 Network Musical Performance

Continual advances in networking technology have led to a virtual collapse of geographical distances. As the field of Computer Supported Collaborative Work emerged to address many of the ensuing social and technological effects, the notion of people being apart yet feeling together has become quite commonplace. Remote collaboration over a network, however, is not a task without its share of challenges for tightly coupled interactive activities, the most glaring of which is arguably latency. Put simply, there are restrictions on the transmission speed of data: while optical fibers that operate at 99.7% of the speed of light have recently been developed [151], existing fiber optics networks can, under the most desirable of conditions, only reach up to 70% [38]. This places the theoretical round trip-time (RTT) between New York and San Francisco at approximately 44 ms, and this figure does not take compression, encoding and decoding, existing traffic on the network or transmission error checking into account. As researchers began to investigate transmission protocols that could address these issues, many turned to music as a testbed for studying the system requirements for synchronous collaborative activities over distance: musical performance is a temporally exacting activity demanding multiple levels of intercommunication amongst the players and, therefore, it is quite suitable for testing stringent network requirements [47]. As far back as 1998, audio applications specifically designed for next-generation networks were initiated to examine the effects of latency and jitter on long-haul uni- and bidirectional data flow. On the other hand, as exemplified through John Cage's 1951 "Imaginary Landscape No. 4", it has become quite common for artists to take "cutting edge" technology and use it to maximize the aesthetic and conceptual value of their work [6]. Thus, it was not long before musicians began to make use of the audio applications described above to experiment with sound over IP from a creative standpoint. The network was no longer limited to being a platform for the unilateral distribution of digital content, but began to act as a medium for high-quality bidirectional musical interaction, propelling the field of distributed performance, more formally referred to as network musical performance [39]. However, performers were quickly confronted with the reality of time delays and quality losses. The unidirectional latency required to achieve synchronous play must be lower than what is known as the one-way ensemble performance threshold (EPT) of only 25 ms [169], a condition that is extremely hard to achieve even under the most ideal of network conditions. To cope with this drawback, a number of artists turned towards exploring alternative approaches for performance over the network. In turn, many began questioning the merit of aiming for synchronous presence with remote participants, choosing instead to investigate the implications of remoteness. Renaud, for instance, considers latency to be a musical feature in its own right that can be used as a specific compositional tool [159]. Similarly, Tanaka describes the instinctive reaction to reduce delay in network performance as a "misplaced motivation", explaining that accepting latency can lead to the creation of music that is idiomatic to the medium [178]. Furthermore, the author likens the network's temporal characteristics to those of any other physical acoustic space, saying: "Seen in this light, it was the same as when composers consider the acoustical characteristic of a concert space in which their work might be performed. Composers of sacred music in the Medieval era were writing for reverberant cathedral architectures. They were fully aware of this, even taking advantage of the long reverberation times to "hide" secular melodies within the long, slowly moving lines of the *cantus firmus*" [181]. Schroeder and Rebelo also encourage us to shift the emphasis of “being there” towards a greater exploration of “being apart”, and further claim that musicians, digital artists and performers are moving towards embracing what computer programmers consider “problematic”, “disturbing” or “irritating”, nd developing strategies for addressing loops, latencies and disruptions that can be characteristic of a typical network [168]. Thus, we argue that there is strong merit to be found in embracing the network as a performance space, with all its idiosyncrasies and their implications.

2.2.1 Embracing the Network as a Performance Environment

Schroeder and Rebelo claim that the network is no longer merely a channel for communication and exchange, but rather a “place in its own right, a space for being, a locus for dwelling” [167]. As exemplified through online virtual worlds and social networking sites, the authors continue, we are no longer onlookers, but active participants *in* the network. This notion begets some interesting implications. The history of music is intrinsically linked to places and societies. Consider, for instance, the introduction of the recording studio. Not only did it lead to a significant shift in our definition of performance, it also paved the way for musical practices that “depend on and are ultimately entangled in, [sic] the studio as a musical environment” [168]. Thus, NMP, many authors argue, should be granted the same considerations, and regarded as both an acoustic and social medium. In a way, the most exciting prospects for NMP lie not in emulating the traditional stage, but in using the network to explore new types of performance and purpose-created music [159]. Many performance environments have been developed to explore the implications of regarding the network as a milieu. Renaud et al., for instance, distributed an instrumental sextet across three sites, each with distinct acoustic characteristics. The authors’ goal was to explore the superimposition of acoustic spaces that is implicit in two- or three-way networks, much in the same way that standard ensemble relies on a common acoustic space between performer and audience [159]. This bears resemblance to the virtual microphone control system, or ViMic, which can project distributed musicians into a “shared virtual acoustic space”. The sound of the musicians’ instruments at one end is captured using spot microphones, then spatially projected at a remote end using an array of loudspeakers [26]. ViMic was an integral component of the Tele-Colonization performance that took place at McGill University in 2007. Musicians at McGill's Tanna Schulich Hall were joined by ensembles at the Rensselaer Polytechnic Institute (RPI) in New York, Stanford University in California and KAIST University in South Korea. The Tele-Colonization performance saw its participants inhabit not only the virtual acoustic space created by ViMic, but also a visual one, whereby still images from each location were incorporated and gradually unveiled at McGill in proportion with the musicians' movements. In addition, audience members at McGill had the option of switching to the acoustic environment at RPI using wireless headphones [25]. Similarly, Barbosa's Public Sound Project was designed to go beyond most common paradigms of NMP, where the network is merely used as a communication channel, and to provide an on-line public performance space where people could choose to participate (either as performers manipulating sound objects, or as members of audience) in on-going collaborative sonic events. According to the author, it is the Internet's essence to provide permanent connectivity. Therefore, a public Internet event should go on permanently, and the audience and performers should be free to join and leave as they see fit [6]. This notion of permanent connectivity also brings to mind the Global String installation [182]. Designed by Tanaka and Bongers as "a musical string (like the string of a violin or guitar) that spans a large geographical distance", Global String consists of two large steel cables, physically separated from each other, but connected as one through a virtual string on the network. Plucking one of the cables leads the other to resonate, both physically and acoustically. As a "musical instrument that exists in the mixed realities of acoustical and network space", Global String was specifically designed to explore the implications of performing across the network. The authors wanted to forgo the goal of "seamless remote communication" typically sought by videoconferencing applications, and instead use Global String to "create an awareness of the network as a visceral entity" that often behaves less than ideally. Audio quality at one end was meant to reflect the distant nature of the remote side, and the conditions of the network carrying the signal across. In addition, like Barbosa's Public Sound Project, Global String is a continuous installation: the string is always present, vibrating, awaiting a user, reflecting the "temporally imprecise nature of the network". Another example is Constellations, a gallery installation that allows visitors to navigate a "spatial acoustical network". Excerpts by different composers are represented by an on-screen universe of planets with which visitors can interact. When a planet is chosen, its accompanying sound is streamed from one of five computers and resonates in the gallery, thereby allowing visitors to effectively navigate through the network space [180]. That being said, embracing the network as a valid performance environment implies confronting the reality its inherent latency, and the effect this has over any musical activities. As a result, taxonomies for NMP tend to classify existing systems in terms of the latencies they exhibit, as discussed in the following section.

2.2.2 Classification of systems for Network Musical Performance

Carôt et al. have identified three design philosophies generally adopted by creators of systems for distributed musical performance [39]:

1. **Realistic Jam Approach (RJA)**: the goal is to enable geographically displaced musicians to feel as though they are playing in the same space.
2. **Remote Recording Approach (RRA)**: this approach involves producing music by using the Internet as a medium for remote recording sessions.
3. **Latency Accepting Approach (LAA)**: the Internet is a decentralized and space independent medium, and thus network delays of more than 200 ms are common and perfectly acceptable.

In this section, we provide an overview of existing systems that exemplify these approaches, and discuss the strengths and weaknesses of each.

Realistic Jam Approach (RJA)

Long before the proliferation of the Internet, distributed musicians had begun taking advantage of available communication technologies to collaborate successfully with one another. In fact, as far back as 1975, Galloway and Rabinowitz used satellite transmissions to network artists performing dance and music scores. For most of the two decades that followed, satellite links were the only means of connecting remote musicians, with the exception of The Hub, whose members used telephone lines connected to a modem to transmit Musical Instrument Digital Interface (MIDI) messages between musicians in 1986. In 1993, Schooler et al. were among the first who used the Internet to synchronize, at one location, three real-time streams of music transmitted from different hosts, albeit at a significant delay of 200 ms [49, 166].

In the following years, Cooperstock et al. set to create the "Recording Studio that Spanned a Continent", where the remote end would be able to mix the received audio signal. In 2000, this culminated in a demonstration during which recording engineers at the University of Southern California were able to mix the 12 channels of uncompressed pulse-code modulated audio streamed from a jazz group performing at McGill. The event marked the first time that live audio sampled at 24-bit 96 kHz was successfully streamed over the Internet [52]. It was also in 2000 that the Internet2 framework was first used in a distributed performance context as part of the World's first Remote Barbershop Quartet. Each of the four singers was in a different location, and although they could not see or hear each other, their efforts were coordinated by a conductor and a mixer who, along with an audience, were present at a fifth location [48].

Finally, in 2001, SoundWIRE (Sound Waves Over the Internet From Real-Time Echoes) was used to organize the first successful two-way musical collaboration over the Internet. SoundWIRE is a Transmission Control Protocol (TCP) based framework originally designed to explore the use of audio as a network measurement tool, through the use sonar-like pings that aurally display the quality of a bidirectional connection. Chafe was able to use SoundWIRE to stream high quality audio bidirectionally between a pianist at Stanford University, California and a cellist at the Internet2 headquarters in Armonk, New York. Although there was very little signal loss, the acoustic latency of 125 ms was "on the 'hairy edge' for an unencumbered performance" [169]. Nonetheless, despite the noticeable delay, Chafe reports that musicians were able to "catch-up" during the pauses [44]. A much lower delay was experienced when musicians from Stanford University and McGill were joined in a cross-continental jazz session in 2002 using the Ultra-Videoconferencing system developed at McGill. Surround sound and full-screen video were streamed bidirectionally over a dedicated communication line. Although the system was unable to achieve a one-way audio delay much below 50 ms, the musicians involved reported feeling as though they were present on the same stage [49].

A number of standalone software applications for audio streaming were later created with accessibility to the average musician in mind. For instance, the readily available Soundjack software, developed by Carôt in 2005, can directly access the soundcard buffer and send audio data via User Datagram Protocol (UDP) [37]. Soundjack was successfully used in a number of distributed performances that attempted to replicate the co-present condition all across Europe. However, the final latency achieved was dependent on the physical distance between sender and receiver, the type of routing between them, and network capacities and conditions. Similarly, eJamming is a commercial software available online that promises to "enable musicians to play together in real-time as if in the same room even if they are far from one another" [35]. It differs from Soundjack in using MIDI data rather than audio, thereby greatly reducing the bandwidth requirements. In addition, data is transmitted only when an event is triggered by the user. eJamming deals with delays in two ways: first, a data package arriving after a time threshold (pre-defined by the user) is discarded. Second, "delayed feedback" is used for any sessions where latency exceeds the ensemble performance threshold of 25 ms, meaning that one's own instrument can be delayed by an adjustable amount in order to get it closer to the incoming sound. Unfortunately, both strategies can have undesirable effects on performance, with the former leading to missing notes and the latter causing unnatural feedback between a musician and his instrument[39]. As another example, the JamSpace system was designed to encourage novices to play music together anonymously over a network, by combining pressure sensitive pads mapped to percussion instruments with a simple software interface allowing users to set various performance parameters and connect with one another [87]. Finally, Alexandraki and Akoumianakis introduced DIAMOUSES, a system they described as an "open framework that aims to enable a wide range of applications and services for intuitive music expression and collaborative performance among geographically separated musicians" [4]. In essence, DIAMOUSES is a generic platform that can support a variety of synchronous and asynchronous distributed musical activities, such as rehearsals, live performances and lessons, by allowing its users to select between peer-to-peer or star network topologies according to their needs. The framework also offers an open and reusable application programming interface, or API, to facilitate integration with other existing toolkits.

We should note that alternatives to the "universal" value offered by the commonly used ensemble performance threshold have also been explored. For instance, during a study conducted by Carôt et al., five drummers were asked to perform separately with the same bassist, all of whom were professionals. The performance speed varied between 60 and 160 beats per minute, and the delay between each of the two players was increased in 5 ms increments, until either of them began to feel uncomfortable or slowed down. The authors found that each player's results varied "in such an extreme way that it is not possible to define a general valid threshold". Instead, they propose defining an "individual acceptance value" that depend on each player's "rhythmical attitude", or what the authors define as the Personal Beat Shift Range (PBSR). Carôt et al. also introduce the notion of the ensemble delay acceptance limit (EDAL), a value that must be determined separately for every dedicated test setup [40].

Interestingly, several experiments with systems developed under the RJA have shown that musicians will develop their own strategies for dealing with large delays when they do occur. For example, the Distributed Immersive Performance (DIP) system, designed to be a seamless environment for remote and synchronous musical collaboration, was used for a series of user-centered experiments to assess the psychophysical effects of latency systematically on remote collaborative musical performance [47]. In their studies, Chew et al. enlisted the help of the Tosheff Piano Duo, two expert users, to perform a rhythmically demanding piece under various conditions of auditory delay. The tolerable threshold for auditory delay was found to be in the 50-75 ms range. Around the range's upper limit, the performers struggled to keep time. Surprisingly, however, tempo variability decreased when the delay increased to 150 ms. The authors hypothesize that greater tempo variability was observed around the usability threshold because, at that point, the musicians began exploring new strategies to compensate for the delay. Beyond the threshold, however, the latency is unacceptable and the players revert to their *modus operandi*, adopting more stable strategies that do not differ greatly from the practiced norm. Similarly, in October 2001, real-time improvised duets between musicians in Dresden and Barcelona took place as part of the Networkshop festival. Even though the experienced latency hovered around 100 ms, the musicians had the freedom to increase the delay as they saw fit. Perhaps surprisingly, they reported having a "very good feeling of playability" as a result. In turn, their strategies greatly affected the resulting soundscape, leading to a performance that can be considered impossible outside the network [6]. As a final example, Renaud and Rebelo held a real-time, three-site tele-concert between Queen's University (Belfast), Stanford, and the University of Washington. To overcome latency, the musicians began adapting a "leader-follower" strategy, which allows for a selected leader in one site to set the tempo for all others. The performance helped uncover that musicians can easily adapt to new types of listening situations, particularly if they remain relatively stable [159]. Interestingly, these examples illustrate that increasing latencies can drive musicians to transform their overall approach into a "latency-accepting" one, a philosophy further discussed in Section 2.2.2 below.

Remote Recording Approach (RRA)

Propelled by popularisation of e-mail, many of the early and experimental web-based distributed performance systems of the 1990s were founded on the Remote Recording Approach. One of the first such systems was Craig Latta's NetJam, which allowed a community of users to collaborate asynchronously by exchanging MIDI files through e-mail [119]. In 1996, the introduction of Steinberg's Virtual Studio Technology (VST), an interface that integrates audio synthesizers and effect plugins with audio editors and hard-disk recording systems, served as a major catalyst in the development of next-generation remote recording applications. For instance, the VSTunnel Plug-In, which was designed to be used like an insert effect in a VST compatible sequencer's master out channel, allows users to start or join other sessions. A session can be made private, meaning that its creator can distribute it as she sees fit (by e-mail, for instance), or public, which adds the session to a public list accessible by other users. When a user joins a pre-existing session, its contents are analyzed by the VSTunnel plug-in. Local changes are then recognized, compressed and transmitted to the other participants in the session. Such changes can subsequently be mixed into the audio output. In this manner, within a local session, every user is able to listen to his own mix, as well as those of the other participants, and adjust each as he sees fit [136]. The Digital Music Link (DML), another plugin for VST, was designed to promote asynchronous collaboration amongst remote musicians. User A, acting as a "performer" and User B, acting as a "recorder", log on to the DML and choose a session. User B assigns a track in his production to User A, then starts the recording process. User A receives the mix and adds in her own track, as though present in a recording booth in User B's studio. Each recorded sample is given a timestamp by the DML to facilitate sorting. User B's playback does not start until User A's data is fully received, so as to maintain the recording booth and studio analogy [163]. Another example of the Remote Recording Approach is the Internet Sound Exchange (ISX) program, an application for computer music composition, performance and improvisation for Internet2 [94]. It allows many hosts to send algorithmically controlled mixes of sound samples to each other. Before a performance, musicians on each host must create their sound sources and upload them. The sounds are then processed, stored and distributed among a number of machines connected to one another over a network. Later, they can be accessed, manipulated and changed by other musicians, leading to improvisations that enable performers to "interact musically as if they were in the same room". However, perhaps the best known example of a system designed according to the Remote Recording Approach is Sergi Jordá's Faust Music On-Line (FMOL). The primary goal behind FMOL's development was to introduce the practice of experimental electronic music to newcomers, while also satisfying more advanced electronic musicians. Users are able to create their own compositions using the system's graphical user interface (GUI) before uploading them as small proprietary score files in a relational database. Other users can then easily access the database, and further manipulate the score files before uploading the revised versions back to the database. What made FMOL stand out was its GUI, where all sound manipulations were represented visually in a manner analogous to the strumming of guitar strings. This made the system fairly easy to learn for novices [105]. As of FMOL 3.0, introduced in 2003, musicians were finally able to use the system for real-time concurrent "net-jamming".

It is important to note, however, that while Remote-Recording Approach systems offer effective solutions for distributed collaboration, they are very much designed with the notion of the traditional studio, rather than the stage, in mind. As a result, such systems revert to a dated view of the network as a mere channel for the exchange of information and, in a sense, constrain what Barbosa refers to as "the potential of what the Internet can offer as a medium for artistic expression" [6].

Latency Accepting Approach (LAA)

According to Tanaka, "Music exists in space, in acoustical contexts, in the environment that it is played in. If music is made on networks, the network infrastructure becomes the space the music occupies. The time characteristic of that infrastructure defines the musical quality of that medium. Network transmission latency thus becomes the acoustic of the network, to be respected and exploited just as one does when composing for specific physical spaces" [181]. Taking this view to heart, a number of artists began developing performances that explored the delays and disruptions inherent to the network, leading to emergence of the Latency-Accepting Approach. In the mid-90s, for instance, the Norwegian art collective "Motherboard" recognized the artistic potential in low-bandwidth transmission, as exemplified, for instance, through pixellated images and choppy sounds. Their 1995 work "M@ggie's Love Bytes", for instance, sought to exploit the asynchronicities typical of network technology [58]. Similarly, the duo "l a u t" performed "A Synk" in 2005, a piece meant to explore the improvisational content between two groups of musicians located in different places and connected by a low-bandwidth chat link. The performance was driven by and shaped by the limitations of bandwidth, unpredictable delays and interruptions [156]. As another example, Chris Chafe used SoundWIRE (described earlier in Section 2.2.2) to create the “Ping” installation, which functioned as a sonar-like detector whose echoes sound out the paths traversed by data flowing on the Internet [44]. At any given moment, several sites were concurrently active, and the tones created by Ping made audible the time lag that occurred while information moved from one site to another. Visitors to the installation could expand or change the list of available sites, and influence the types of sounds produced by choosing different instrument projections, musical scales, and speaker configurations [43]. The “Gigapop Ritual” is another instance of the Latency-Accepting Approach. In fact, the performance was created specifically to apply Tanaka’s view that “[l]atency is the acoustics of the Internet”. Musicians located at McGill University engaged in a network performance with musicians at Princeton [107] as part of the 2003 Conference on New Instruments for Musical Expression (NIME). High-bandwidth, bi-directional real-time audio, video and controller data was streamed during the collaborative event, which involved new digital musical instruments and traditional Indian ones. The goal of the performance was not necessarily to explore the effects of latency on the performers, but rather to allow them to experiment with different rhythms and soundscapes through free-form improvisation with one another. An important aspect of the performance was to explore multiplayer digital controllers by networking musicians at the geographically different sites. This type of system is in fact known as an Interconnected Musical Network, a topic that will be discussed in Section 2.3.1 below. One more example of the Latency-Accepting Approach is the Master Cue Generator (MCG), a system designed to provide musicians with various cues in an effort to help them understand and cope with the effects of large latencies. The MCG allows a “central” node on a network to act as a server, sending three types of cues to other connected nodes: temporal, behavioural and notational. Temporal cues send out information such as the length of a cue, a warning that a cue is about to finish, or how much time a given node is in control of the improvisation until it delegates control to another node, thereby changing the network’s topology. Behavioural cues are sent with a certain scenario attached to them, which can include the triggering of a waveform, or a suggestion that a given node should only play certain notes. Notational cues can include the visualization of a waveform from each site, the display of the cue number, a countdown or dynamic shapes that can be activated by various factors in the performance. In order to deal with latency, the MCG provides two approaches that are defined as synchronous or asynchronous. For synchronous interactions, latency is added to all cues being sent such that all nodes experience the same delay. In the more interesting asynchronous approach, all nodes experience latency as is, which leads to the generation of rhythmical patterns created by the network itself [158].

A number of software tools have also been made available to the average musician wishing to experiment with the Latency-Accepting Approach. For example, the Novel Intervallic Jamming Architecture (Ninjam) tries to establish a jamming environment under the assumption that network latency prevents true real-time synchronization of the participating musicians. Users receive each other's output with the delay of at least one measure, which Ninjam's creators call "faketime". The goal is to put emphasis on musical experimentation and expression rather than synchronicity [75]. Another example is Quintet.net, an interactive distributed performance environment that enables performers at up to five locations to play music over the Internet under the control of a central server acting as a "conductor". Musicians send control streams to the server using either a pitch-tracker, MIDI signals or the computer keyboard. The server then copies and processes the streams, before sending them back to the musicians, as well as any interested listeners [88, 39].

While the taxonomy for latency approaches offered by Carôt et al. (RJA, RRA and LAA) allows us to categorize the vast majority of NMP systems, it can be characterized by an almost exclusive focus on the auditory and temporal properties of distributed performance. However, as Braasch et al. explain, "[v]isual cues can also be instrumental to negotiating the solo order in improvised music, or enabling social exchanges, such as signalizing to someone that her solo was well received"[27]. Thus, in the following section, we discuss paradigms for visual communication between remote players that were specifically designed to capitalize on the unique nature of network performance.

2.2.3 Visual Representations

Playing music is a multimodal experience: a musician looks at his instrument, other musicians and his audience; he listens intently to all sounds produced around him; he relies on haptic feedback from his instrument for guidance; and, all the while, he communicates with other musicians, using both words and gestures. Unfortunately, such interactions can become a real issue in network musical performance, as many cues are completely removed from context. In fact, The Master Cue Generator described above was designed specifically to help re-introduce behavioural cues that, while often taken for granted in the co-present setting, can disappear in distributed performance, much to the players' detriment [157, 158]. Nonetheless, when it comes to visual communication in particular, many distributed performance systems continue to take their lead from traditional videoconferencing, offering full-frontal video as the only solution. We find that this approach does not take into account the network's unique characteristics. In an effort to resolve visual communication during NMP, many researchers believe it is first imperative to better examine the implications of being *in the network*. For instance, Schroeder notes that "in the same way that you cannot stare at the network straight in the eye, that you can never directly confront the network, for it is always somewhere else from wherever you may be looking, performers never stare at other players"[168]. In traditional performance, although musicians communicate with each other through various cues and body language, they do not stare at each other directly, and do not require a full-frontal view of one another. Instead, each musician experiences only "fragments" of the whole performance environment, through glances and peripheral vision. In fact, only the audience is able to experience the ensemble visually as a cohesive whole. On this notion, Schroeder further proposes that "equipping performers with a full-frontal visual perspective of remote players fails to address the intricacies of performative interaction, which are rooted in interpretation rather than in communication, in the fluid rather than the representational" [167]. Putting this philosophy to the test, Schroeder et al. started the "Apart Project" in 2007 as a study on various novel network scenarios. Three musicians were asked to perform two very different songs over a network: one requiring a high degree of rhythmic synchronization, and another that allowed for rhythmic improvisation based on the realization of a set of instructions. The Apart Project was divided into various scenarios that allowed the authors to explore a number of conditions, including the use of avatars and standard video conferencing technologies. The avatars were designed as close-up and detailed yet abstract renditions of performance gestures. They emerged from the understanding that one never sees one's own body as optically complete, but rather as fragmented. Upon examination of video footage collected during the experiment, the authors found that the avatars were not particularly helpful when playing the first piece, because musicians could not focus on the score and a computer screen at the same time. However, for the second piece, which required an acute type of listening, the performers constantly looked at the 3D avatars. In fact, post-test questionnaires revealed that they enjoyed looking at the avatars "as a means for visual interaction and potentially for enhancing social interaction." [167]. Interestingly, as soon as an improvised section would start, the performers would turn towards the screens on which the avatars were displayed, using the computer-generated graphics as a way of interacting with each other. When iChat was used to stream full-body video capture amongst the musicians, there were remarkably few glances towards the screen during both pieces, supporting Schroeder's theory that performers do not need to stare at one other directly and constantly. The contrast in behaviour observed when full-frontal video was used instead of avatars implies that musicians have quite an abstract reading of each other's presence. By being one level removed from full-body representation, the 3D graphics that constituted the avatars required interpretation, much in the same way that a performer's glance demands in traditional performance [167]. As another example, Konstantas et al. also reported an interesting experience with regards to the use of shared video in a distributed context setting. In 2001, they developed the Distributed Music Rehearsal project, an Asynchronous Transfer Mode (ATM) based system that allows small groups of musicians to rehearse with a remote conductor. In contrast with the Apart Project, where the goal was simply to facilitate collaboration between remote musicians, participants in the Distributed Music Project were expected to pay a greater level of attention to the conductor, as is common in rehearsal settings. The authors indicated, however, that musicians found giving "continuous attention to the projected video" of the conductor to be tiring. In addition, musicians were unable to determine where the conductor was pointing due to the flatness of the video monitors [117]. Like Schroeder, Kapur et al. have experimented with the use of specialized graphics, in addition to full-frontal video, as part of their work on distributed performance. More specifically, the authors developed the *veldt* software, described as a "real-time networked visual feedback software" that can trigger arbitrary text, images, videos or geometric models in response to MIDI events [106]. Mappings are flexible and can be set by the musicians prior to a performance. For instance, when *veldt* was used as part of the Gigapop ritual (described in Section 2.2.2), remote players of the Electronic Dholak (EDholak), a multiplayer networked percussion controller based on the Indian Dholak, were allowed to interact with one another through a sculptural metaphor. The events they generated by striking the EDholak were "dynamically mapped to a series of geometric operations that generated, deleted, deformed or detached" elements of a visual artefact. Not only did the metaphor render the EDholak players' actions visible to one another, it also encouraged them to interact on a new level through their collaboration to shape the artifact itself. Animated graphics and abstractions were also a pivotal aspect of the Global Visual Music (GVM) project, which evolved to investigate "sensory connections through physical action, moving images and improvised music" [173]. The project comprised of a series of live distributed performances, augmented with video and computer animations created by Sorensen. Steiger composed structures for the improvising musicians, and Puckette provided software that generated the necessary control streams between the music and animations. This software would later evolve into Pure Data, described later in Section 5.1. Sorensen describes the goal of the project as exploring the "abstraction of connection", by bridging the audience's visual and auditory senses in a manner that is more complex and challenging than the traditional sonification of visual elements, or visualization of sonic elements. Global String, described earlier in Section 2.2.1, is another example of a distributed music installation that utilized special graphics: in addition to shared video, a visualization of the audio signals consisting of "a real-time amplitude envelope and FFT frequency spectrum of the sound from the two sites", as well as a visualization of the network conditions using readouts from the Unix traceroute command were used at each end of the string [182]. It is also worth noting that Tanaka and Bongers did not simply rely on traditional videoconference techniques, but designed their shared video setup with the transduction of data from the physical space to the network, and back to the physical space, in mind. As such, cameras and monitors were physically arranged in a manner that faithfully recreated "the sense of playing along one string": their orientation was chosen to preserve eye-to-eye contact between players, and create the illusion that both ends of the steel cable were in fact physically connected. While the authors did not comment on the success of their special visualizations, they did provide some notes regarding the shared video, stating that while "[v]isual contact with the remote player complements the audio connection. . . sound quality is more crucial than image quality" in any musical project. This is in direct agreement with the views of Kapur et al., who state that the importance of stable audio supersedes all else in NMP, explaining that "[w]e can tolerate a dropped or repeated video frame now and then, but not choppy audio" [106]. Interestingly, during performances with the Distributed Music Rehearsal system, Konstantas et al. observed that "musicians and the conductor preferred to have a smaller delay of the audio. . . than to synchronize it with the video which had a much longer delay" [117]. In contrast, however, Tanaka noted through his extensive work on network musical performance that "[t]he image component's contribution was effectively nullified unless the image was synchronized with the audio" [182].

In conclusion, there are a wide variety of approaches to sharing visual information in distributed performance, ranging from minimal abstract animations to high-quality, full-frontal video, and even combinations of the two. Such visual representations can also be seen as analogous to the surrogate gestures explored by Cornelius et al. within the context of CSCW, and described earlier in Section 2.1. What is of interest to us, however, is whether overall performance quality can be improved by sacrificing video in favour of low-bandwidth paradigms for visual communication, a question we further explored in this dissertation.

2.3 Additional Topics of Relevance

Given that our work seeks to extend the notion distributed performance, it shares a variety of common traits with a number of specialized research areas beyond network musical performance. In this section, we discuss these various domains in an effort to better situate our research within existing applications.

2.3.1 Interconnected Musical Networks

It comes as no surprise that the act of distributing performance over a network would have a strong impact on the nature and level of communication between remote musicians. Renaud, for instance, explains that "[i]nteraction is a real issue in network performance systems as natural visual or sensory cues, such as breathing and gesture, are completely removed from context" [39]. To this, Kapur adds that "[w]aiting backstage to go on, and important aspects of socialisation after a performance, are not the same over a network", leading to a "loss of society within the band" [106]. As we wanted our system to be driven largely by player-to-player interaction (a criteria we discuss later in Section 3.3), we were particularly interested in music technology applications that focus on increasing the level of interplay between musicians. One such research area is that of Interconnected Musical Networks (IMNs). A term coined by Weinberg in 2002, Interconnected Musical Networks are live performances where players can influence, share and shape each other's music in real-time, and can even be seen as an example of the Collaborative Digital Musical Interaction paradigm, described earlier in Section 2.1. Naturally, traditional performance can, to a certain extent, be considered a form of IMN, as music-playing is a highly interdependent art form. Nonetheless, while co-present musicians can influence each other a great deal, the level of control over this influence is rather limited. For example, a soloist can steer her collaborators towards a musical idea in which she is interested, but this type of influence is more of a suggestion. She has no direct control over the other musicians' instruments, and there is no guarantee that they will consent to her desire

[191]. As the introduction of new musical interfaces facilitated the construction of electronic communication channels between instruments, musicians became able to take a fully active role in determining not only their own musical output, but also that of their peers.

Interconnected Musical Networks differ most notably from NMP systems in the fact that they do not necessitate participants to be apart, and can in fact be used in a shared space. In fact, many consider John Cage's 1951 "Imaginary Landscape" to be the first example of an IMN. Two performers were assigned to each of the 24 radio transistors used, one as a "frequency-dial player" and the other as a "volume-dial player". The score indicated the exact tuning for each radio at any given time, but without any foreknowledge of what might be broadcast on a station, or if one even existed at the specified frequency. The volume player could then manipulate his corresponding frequency player's output, deciding whether it should be a slight whisper or a screeching solo [191]. When the commercialization of personal computers began in the 1970s, the League of Automatic Music Composers became the first group to write interdependent computer compositions. Dubbing the new genre "Network Computer Music", the group set up a three-node network, mapping frequencies from one computer to generate notes in another, or mapping intervals from one composition to control rests and rhythmic patterns in another. The League of Automatic Music Composers evolved into The Hub in 1986, as its members improved their performance by using MIDI data (as described earlier in Section 2.2.2) exchanged through a central computer (or hub) rather than ad-hoc wired connections. Other examples of IMNs include Duckworth's "Cathedral", which in fact was the first piece composed specifically for the Web in 1997. Using a Java applet, participants could trigger sounds by clicking on nodes hidden in the screen. Although the original sounds were composed by Duckworth, players could contribute their own sounds to the mix. Seeing as there was no connection between the players, the system could support any number of users [62]. Another example is "Variations for WWW", an application introduced by Yamagishi in 1998. The goal of the project was promoting "interactivity as opposed to unilaterality" and "sharing as opposed to monopolizing" [194]. Remote users could access a MAX patch connected to the Internet and manipulate parameters that were subsequently sent to a MIDI synthesizer. The resulting output was then transmitted back to the participant. Users could play the combined manipulations of all the other users, and modify their own contribution in response. Similar to this example is the Palette, an online system that not only allowed participants to share music in the form of MIDI events, but also to control the “style and “energy” of content uploaded by others [196].

Interconnected Musical Networks can be classified by their topologies or architectures. They can be centralized or decentralized, and symmetric or asymmetric. Centralized networks allow players to interact through instruments or controllers that do not have a direct influence on each other, whereas decentralized networks enable musicians to interact directly with one another (see Figure 2.2). Centralized networks can, in turn, be synchronous, where players can manipulate the music of their peers while it is being played, or sequential, where each player must submit their musical material before it can be affected by a peer. Symmetric networks are ones where all players have the same level of control. Asymmetric networks, on the other hand, allow connections only in certain directions and only among certain nodes (see Figure 2.3). Asymmetric networks may also assign a weight to each player’s influence, giving some the ability to effect more change than others.

Despite the promise to enhance the level and quality of interaction between musicians, participation in early IMNs was not typically a simple process. As seen with The League of Automatic Composers, for instance, the majority of interdependent connections between players were based on low-level elements, requiring participants to possess specialized musical skills and technical knowledge in order to partake meaningfully in the process. As a result, the cognitive load required from performers to manage such interactions prevented them from further exploring the social and expressive aspects of the network. Recalling one of The Hub’s earlier performances, member Gresham-Lancaster commented: “The technology was so complex that we were unable to reach a satisfactory point of expressivity” [82]. Furthermore, the interactions proved to be too complex for audiences to understand. In fact, Weinberg believes that such issues are not uncommon even with modern IMNs, some of which continue to focus on complex connections that force performers and audiences

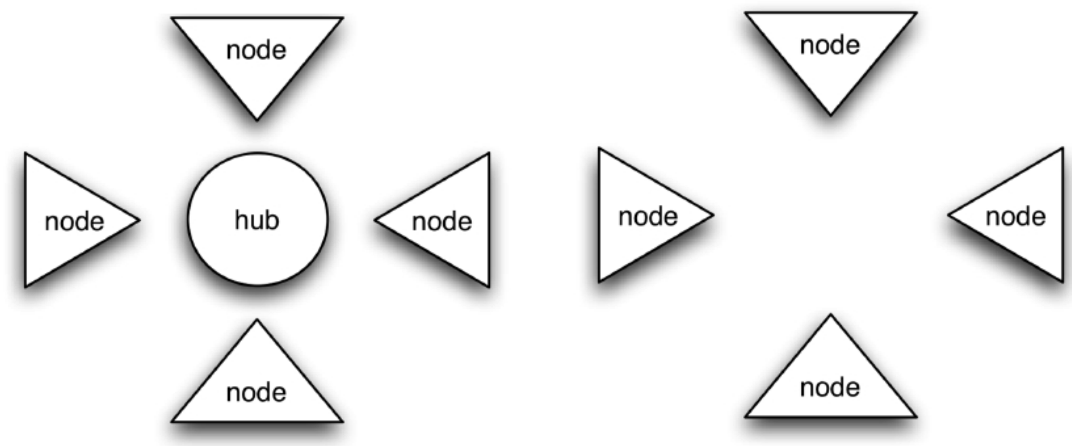

Fig. 2.2: Centralized (left) and decentralized (right) Interconnected Musical Networks, adapted from reference [191].

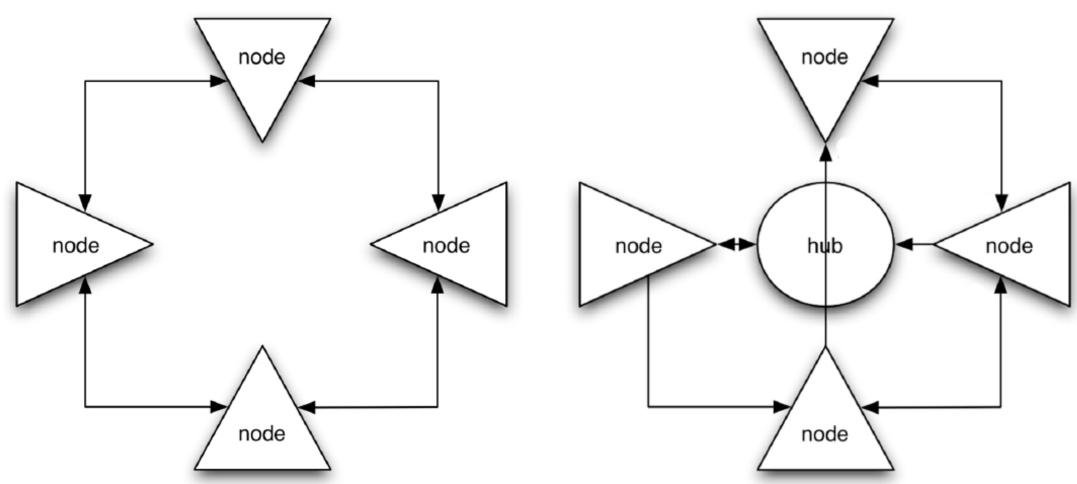

Fig. 2.3: Symmetric (left) and asymmetric (right) Interconnected Musical Networks, adapted from reference [191].

to focus on “low-level analytical elements” rather than “the expressive and social aspects of the network”. A solution he proposes is to allow IMNs to take on the form of expressive, gesture-based physical instruments, thereby making the overall experience “more intuitive and accessible for novices, wide audiences, and even children” [190]. This idea bears a strong resemblance to the philosophy behind interactive installations, another type of interfaces that also encourages interplay between multiple users, all while remaining completely accessible, a notion we expand on in the following section.

2.3.2 Interactive Installations

Given our interest in exploring specialized forms of interaction, we turned to the existing body of work on interactive installations for further guidance. We use the term interactive installations (IIs) to denote works that are commonly referred to in the literature as “interactive sound installations”, “interactive art installations” or simply “interactive art”. Similar to Interconnected Musical Networks, interactive installations are an example of highly collaborative interfaces, that should, as Blaine and Fels describe, “cultivate meaningful social interactions and experiences for the players” [18]. In addition, both IMNs and IIs are designed with a focus on overall experience, rather than musical outcome. However, interactive installations also differ from IMNs in a number of ways. First, while the entry fee in terms of musical expertise can vary widely for IMNs, IIs are designed with public accessibility in mind. Ideally, participants should be to walk up to an installation and fully explore it with no prior training or experience. In addition, an interactive installation is typically a vehicle for communicating its creator’s message or intent by means of audience interaction with the work [96]. IMNs, on the other hand, serve as musical instruments for performance or composition that encourage higher levels of interaction between participating users. An example of interactive installations is Iamascope, where a camera captures viewer images and movement that are in turn used by a controlling computer to project corresponding kaleidoscope-like images and creating accompanying music [69]. Absolute 4.5 is another example, where participant presence is determined through floor sensors and used to generate a large grid of colours projected on a screen and a complex soundtrack. In essence, the system's behaviour is mostly determined by audience behaviour, making the performance somewhat unpredictable [17, 63]. The Intelligent Street System further illustrates the accessible nature of IIs: as an alternative to the often undesirable "Muzak" heard in public spaces, it allows users to request changes via mobile text messages. The overall result is to turn visitors of a space from passive consumers to active participants creating their own aural landscape [131]. Similarly, the Control Augmented Adaptive System for Audience Participation (CAASAP) was a project designed to examine a variety of ways in which audience members could make use of mobile phones to become part of the music-making process [177]. Finally, Feldmier et al. created low-cost wireless motion sensors that enabled them to estimate the level of activity of a large-scale crowd. The data could subsequently be used to generate music and lighting effects, thereby allowing members of the crowd to drive the music to which they danced [67].

Interactive installations aim to engage the audience, whose participation is a crucial aspect of the realization of the work itself. According to Candy, "[t]he artist, the technologist, and the audience are all participants in the conceptualization, the construction and the active experiences of the work". The author further explains that such interactive systems are "as varied as the individual people who interact with it" [33]. To this, Bilda adds, "The design process of interactive art systems involves systematically examining audience behaviour starting with what attracts them, what initiates their interactions, following with sustaining their engagement across the overall experience of the work" [16]. He also explains that such requirements make user-centric approaches a common and recommended practice among designers of IIs, a topic we further discuss in Section 2.4 below.

2.3.3 Responsive Environments

Interactive installations bring to mind another area offering rich examples of hands-free, highly specialized interactions: responsive environments. In fact, we consider responsive environments, also sometimes referred to as reactive environments, as the more utilitarian counterparts of interactive installations. However, while users partaking in interactive art are typically aware of the process they become one with, the most defining tenet of a responsive environment was perhaps best described by Elrod, who said such systems should do their job "well enough that the occupants are usually not aware of its presence" [65]. Cooperstock et al. further define reactive environments as spaces where technology, rather than humans, manage the low-level operations of a room [51]. Considered by many to be an extension of ubiquitous computing, responsive environments gained momentum in the 1990s as a solution to "reduce the cognitive load of the user by allowing the system to make context-sensitive reactions in response to the user's conscious actions" [53]. The concept can in fact be traced back to Elrod, who had sought to interconnect Xerox PARC's rich computational infrastructure with a computerized building-management system that could save energy based on office occupancy. Dubbed the "Responsive Office Environments", the system made use of small, low-cost sensors to determine whether a worker was present in her office, and made changes to heating, air conditioning, lighting and desktop appliances accordingly. The Responsive Office Environment was essentially invisible, fulfilling its job while employees went about their daily routines, uninterrupted.

The successful implementation of a responsive environment is highly dependent on effective proximal sensing as the basis for context-sensitive interaction [32]. The Responsive Environments Group at the Massachusetts Institute of Technology (MIT) has investigated this topic extensively. One example of such work is WristQue, a wristband that can detect its user's location and wrist orientation, as well as its environment's lighting, temperature and humidity conditions. While WristQue also doubles as a control interface, allowing users to select and manipulate a variety of controllable devices in the environment, such as light switches and thermostats, through simple gestural interaction, it can also provide each user with personalized automatic control of their environment using a combination of unique identification and location sensing [135]. As another example, Aldrich et al. developed a set of sensor nodes that were deployed across the MIT Media Lab for the purpose of studying the motion patterns of large groups of occupants as a whole. The authors hope that such data can be used to design more effective responsive environments that can, for instance, optimize lighting control in open-floor plans in accordance with foot traffic patterns [3].

Responsive Environments have also proven to be effective solutions for simplifying navigation through complex videoconferencing systems. For instance, Cooperstock et al. created the Reactive Room in response to the frustrations experienced by users interacting with videoconferencing systems. At the time, state-of-the-art videoconferencing rooms typically came equipped with cameras, monitors, VCRs, digital whiteboards and electronic document cameras, all intended to facilitate collaboration amongst geographically displaced workers. However, a session that involved the use of more than one of these tools often proved to be so complicated that a trained expert was required to operate them. The Reactive Room removed this burden by responding to a user's high-level actions instead, letting the technology itself manage the low-level operations between the various pieces of equipment. In a sense, the user interface was made invisible, allowing remote users to concentrate on their collaborative work instead [51]. A more recent, yet similar example can be seen in HomeProxy, a physical proxy prototype that aims to support seamless video communication in the home among distributed family members. Designed to look like a standard home appliance such as a lamp, HomeProxy consists of a slightly bowed rear-projected screen, fabric sides and a wood top. The system was conceived with a "no-touch" interface that utilizes a Microsoft Kinect to detect users and respond to their presence, "waking up" as they approach and "going to sleep" after they leave. Users can subsequently wave hello to begin videoconferencing with a remote family member via Skype, and wave goodbye to end the session [183]. A related example is the Perceptive Presence Lamp, a set of two lamps and cameras that virtually connect two separate locations through awareness information. Each local lamp changes colour depending on the state of a remote occupant who can be absent, working alone at a desk, busy with other occupants, or available for communication [15].

The developers of the Reactive Room believed that the questions they tackled "are not endemic to videoconferencing but apply equally well to other physical environments such as power-plant control rooms, flight decks, and so-called 'smart homes', as well as to software environments such as integrated office suites". In fact, Bongers uses the term "interactivated spaces" when referring to reactive systems, while clarifying that "[t]hese environments come under a variety of labels: hybrid spaces, responsive environments, augmented reality, houses of the future, depending on whether they are developed by groups of artists, architects, and/or interaction researchers" [23]. We find the philosophy behind these systems to be quite suited to musical performance, where technology holds the potential to augment music-making seamlessly with new possibilities, all without distracting musicians from the task of performance itself. In fact, such a notion was explored by Livingstone and Miranda in 2004. The authors developed a novel sonic controller that "regenerates a soundscape dynamically by mapping 'known' gestures to influence diffusion and spatialization of sound objects created from evolving data", and dubbed their system a "responsive sonic environment" [126]. Shortly after, Salter began to explore the use of responsive environments for traditional live performance. The result was Schwelle, a large-scale interactive theatre performance where the rhythm and exerted force of the performers' movements were used to change a musical composition dynamically to "give the impression of a living, breathing room for the spectator" [164]. Overall, regardless of their application, effective responsive environments must be tailored to their users, interpreting their intentions accurately in order to respond to them effectively. After all, an "invisible" system that reacts to a misinterpreted user objective can be rather disastrous. Bongers, for instance, explains that "[w]hen the computer becomes ubiquitous ... misunderstanding also becomes ubiquitous" [24]. Therefore, we argue that responsive environments are best designed through user-centred techniques, a methodology on which we elaborate further in Section 2.4.

2.3.4 Embodied Interaction

As described above, the most significant advantage of responsive environments is their ability to detect and respond to their users' needs without requiring that they detach themselves from the higher level task at hand. Users, in turn, typically exhibit such needs through physical interaction with their surroundings, a notion known as embodied interaction, and further described by Antle et al. as an approach that “involves leveraging users’ natural body movement in direct interaction with spaces and everyday objects to control computational systems” [8]. The idea of embodiment was formally introduced to the HCI community in 2001 by Dourish [59]. While embodiment traces its origin back to phenomenological philosophy [124], which places emphasis on the role of action, experience and perception in meaning making [8], Dourish offers a simple, high-level workable definition as “the property of being manifest in and as part of the world” [60]. “Embodiment”, the author continues, “constitutes the transition from the realm of ideas to the realm of everyday experience... The setting within which the activity unfolds is not merely background, but a fundamental and constitutive component of the activity that takes place.” [60]

In turn, the idea of embodiment leads to some interesting implications when applied to interaction design. As Dourish explains, “[t]he history of HCI can, in many ways, be seen as an ongoing attempt to capitalise on the full range of human skills and abilities. These are not the skills we acquire from training and careful practice, but rather those everyday, natural abilities that most of us take for granted; picking up a ball, not juggling with it” [60]. In addition, Kirsh posits that “we think with our bodies not just our brains” [115]. Thus, if treated as a “fundamental feature of interaction, rather than as a side-effect of interactive system development” [60], carefully designed embodied interactions can potentially capitalize on a broader spectrum of human ability and knowledge [124].

Embodied interaction has proven to be a popular option for the design of non-utilitarian applications, particularly those of an abstract or artistic nature. For instance, Antle et al. designed Springboard, an interactive multimedia environment that allows users to explore the concept of balance within the abstract domain of social justice through embodied interaction [7]. Springboard allows users standing on an input platform to respond to images and sounds representing various social issues by changing their balance to reflect their views on such issues. Another example is Loke and Robertson’s ByStander, an interactive installation intended for public use in a museum or gallery [129]. By sensing the visitors’ patterns of motion and stillness, Bystander responds with a corresponding “spirit-world” of images, texts and sounds drawn from crime scene archives. The authors' goal was to explore, understand and, in turn, leverage their visitors' motion into meaningful and immersive experiences. Furthermore, embodied interaction design has developed strong ties to the study of dance, with a number of HCI researchers exploring this art form as a means of furthering their understanding of movement, and the body's potential as a source of input [115, 128, 130]. Finally, Loke et al. have also examined movement from a gaming perspective by asking participants to play two computer games and subsequently analyzing their motions using existing frameworks for the study movement-based interaction [127].

The idea of embodiment is deeply rooted within the musical context as well, with Godøy et al. describing the well-established links between musical sounds and sound-producing movement as an "embodied understanding of music perception and cognition". Embodied music cognition views the relationship between sound and movement as having its roots in the broader paradigm of embodied cognition, which stipulates that people relate perception to mental stimulations of associated actions [79]. Our work, however, applies the notion of embodied interaction more commonly found in human-computer interaction research, and formalized at the beginning of this section, to the design of musical interfaces. Examples of such embodied interaction within the context of music include the Sound Maker, a room-size interactive audio environment designed by Antle et al. specifically to explore alternatives to desktop-based interactions for electronic music composition. Sound Maker tracks a user's position through a camera, and subsequently maps their location, along with the quantity and quality of their movements, to changes in the pitch, tempo and volume of a percussive audio output [8]. Another example is Corness and Schirphorst's Ariel system, which responds to gestures utilized by musicians during improvisation with simulated breathing sounds [55]. In fact, the authors sought to capitalize on the ability of skilled musicians to exchange, detect and tacitly respond to cues for interpersonal interactions. In turn, the simulated breath generated by Ariel was effectively used "to engage the performer's sense of intuition and empathy while capitalizing on their embodied knowledge of upcoming actions when interacting with autonomous computer systems in performance". Furthermore, Bakker et al. advocate the use of embodied metaphors within the context of musical learning for children [10]. Embodied metaphors are interactions that leverage the notion of embodied schemata, or higher-order cognitive structures that emerge from recurring patterns of bodily or sensori-motor experience. The authors effectively applied this notion through the Moving Sounds Tangibles, a system that allows children to learn abstract sound concepts such as pitch, volume and tempo by manipulating a set of interactive tangibles designed in accordance with various schemata.

Finally, we note that, perhaps unsurprisingly, the study of movement is inherently user-driven. For instance, Bakker et al. consider a "people-centered, iterative approach" to be crucial to successful embodied interaction design [10]. Similarly, in a bid to "foster more meaningful, reflective and satisfying engagement", Loke and Robertson have utilized various aspects of participatory design (a topic further discussed in Section 2.4.2) as part of their study and design of embodied interfaces [128, 129]. We elaborate on the advantages behind such user-driven approaches in the following section.

2.4 User Involvement in Design

2.4.1 User-Centered Design

For years, the flashing "12:00" on VCRs exemplified the curse of innovation created without regard for its user [50]. Nonetheless, while VCR technology itself has become obsolete, complicated, frustrating and unclear interfaces have remained. To many users, such interfaces are simply a fact with which they have learned to live. Naturally, it is not the intention of engineers and systems developers to confuse or aggravate their users. The problem, according to Bongers, is that "[g]enerally computers do not do what the user wants, but what the engineers and designers think the users want, or what the engineers and designers want the users to want" [24]. A developer typically works with a *design model*, based on how he believes a system should behave. He attempts to communicate this model to the user through the *system image*. The user, on the other hand, has developed his own model, the *user model*, through interaction with the system. Problems arise when the system image does not convey the design model well enough to the user, leading him to develop an incorrect mental model: ultimately, the system's behaviour will fail to match his expectations. The mismatch between the user and design models occurs because developers often believe that an ideal interface is one that reflects a system's underlying model, or that the ideal interface for them will also be ideal for the typical user. Norman attributes this problem to a type of "folk psychology" that designers tend to develop, as they project their own rationalizations and beliefs onto the rationalization and beliefs of others [142]. The user, however, generally has no interest in or understanding of a system's inner workings, but is more concerned with completing a particular task using the system [77]. Developers of new musical interfaces are not immune to this phenomenon. Jordá identifies idiosyncrasy as the biggest problem with new musical controllers, stating that many new instruments wind up only being used by their own creators [105]. This view is shared by Orio et al., who also describe the design of NMIs as "marked by an idiosyncratic approach", especially when compared to the design of input devices in HCI [145]. Poepel agrees, and attributes the problem to the fact that the evaluation of NMIs is "often done by the developer or a small number of people" [150]. Such issues tend to occur because developers of NMIs typically see themselves as one and the same as their target users. Their understanding of every aspect of the system at hand, however, prevents them from knowing what may be perceived by another user—even one with the same level of musicianship—as complex. To address these re-occurring usability problems, both within the context of HCI and music technology, a number of designers began turning towards their target users for insight.

The process of systematically involving users throughout a system's design and development cycles was referred to as "user-centered design" (UCD) by Norman, who helped popularize such an approach through his seminal 1986 book on the topic. To summarize the author's view, UCD is the attempt "to ask what the goals and needs of the users are, what tools they need, what kind of tasks they wish to perform, and what methods they prefer to use" [143]. Concurrently, Gould and Lewis devised a concrete UCD methodology by distilling the best known user-centric practices from

HCI research at the time. Their approach was based on three key principles: early focus on users and tasks, empirical measurement and iterative design [81]. Gould et al. used the 1984 Olympic Messaging System (OMS) to simultaneously apply and test the merit of this design methodology [80]. Although the principles proved successful, they provided more of a general guideline for designers, who were then advised to further define and choose the specifics of user involvement in their own work. In fact, when Vredenburg et al. conducted a survey among attendees of the CHI 2000 conference, they uncovered 13 distinct techniques commonly used by UCD practitioners: field studies (including contextual inquiries), user requirements analysis, iterative design, usability evaluation, task analysis, focus groups, formal heuristic evaluations, user interviews, prototypes without user testing, surveys, informal expert reviews, card sorting, and participatory design [188]. This is consistent with Norman's view that pluralism is at the essence of UCD: there is no single best approach that provides a design solution [143]. Ultimately, user-centered design dictates that emphasis should be placed on the user rather than on the technology, and that products should accommodate their users, rather than the other way around.

We note that a more recent trend in the HCI community has seen a shift away from the term "user-centered design", and towards labels such as "human-centered" or "people-centered" perspective instead, due to the connotations they imply. Most notably, Bannon argues that the idea of human-centered computing marks a paradigm shift: rather than explore where users fit within the process of automation, developers should regard design, development and use of software systems as inherently human activities that are fundamental to the computing discipline [12]. In other words, instead of focusing on how specific tools can be designed to help users accomplish specific tasks, the human-centered perspective encourages developers to strive for a better understanding of how people live in the world, and to design systems accordingly. Artefacts, Bannon further argues, are not simply tools, but components of a dialogue between humans and their environment. Given its emphasis on "how people live in the world", rather than the minutiae of their behaviours, we argue that the idea of people-centered design lends itself quite well to experience—rather than task-based evaluation, a topic further explored in Section 2.5.3.

2.4.2 Participatory Design

While Vredenburg's survey, described in the previous section, lists participatory design (PD) as a "commonly used UCD method" [188], many of its proponents view it as a separate and entirely different approach. Similar to user-centered design, PD encompasses guidelines aimed at tailoring end products to their users' needs. However, the major difference between UCD and PD is the nature of the user's involvement in the design process: whereas UCD involves designers calling on users at regular, pre-defined intervals of the design process, PD partners creators and end users as active stakeholders throughout the entire process. Furthermore, while UCD can be regarded as a relatively unidirectional method where prototypes are used as a means of collecting user feedback that is subsequently incorporated as the designer best sees fit, PD involves a two-way exchange of information [138]. Additionally, and somewhat in contrast to UCD, PD not only stems from but continues to be employed in a large number of fields beyond interface design. In fact, originally named "co-operative design", PD emerged in Scandinavia in the 1960s and 1970s as a method of involving members of trade unions, rather than just managers, in the making of decisions affecting workplace conditions. Today, the field, as described by Muller, is "extraordinarily diverse, drawing on fields such as user-centered design, graphic design, software engineering, architecture, public policy, psychology, anthropology, sociology, labor studies, communication studies and political science". The author further describes PD as "the third space in HCI", an "in-between" region where various participants can combine their knowledge and reach new insights that better meet their mutual needs [138]. Similarly, Fowles believes that collaboration between diverse parties can transform "symmetry of ignorance", or a lack of comprehension between designers and users, into "symmetry of knowledge" [74]. Muller proposes a number of strategies for information exchange during the PD process, such as workshops, storytelling and games. One that we found particularly relevant to the nature of our work is low-tech prototyping. Given that novel musical controllers and environments do not naturally lend themselves well to paper prototyping (the approach traditionally employed in the early stages of many HCI applications), we found low-tech prototyping to be an effective method of introducing to end users new technologies as they evolve throughout a project's life cycle. As users are encouraged to think of and form relationships with technologies they may not have been exposed to before, low-tech prototyping promotes a deeper user contribution than conventional paper prototyping [138]. Our adoption of this technique is described later in Section 5.1. Another participatory design method that proved to be of relevance to our system design is that of cooperative prototyping, a topic we further address in Section 8.1.

2.4.3 Music-oriented HCI

As explained earlier in Section 2.4.1, many NMIs fail to find adopters outside the world of academic research. Geiger attributes this problem in part to a lack of established guidelines for the design of NMIs. He explains that since mapping strategies for novel controllers suffer from "missing interface standards and little design experience", a "try-and-error approach" is more often adopted by developers [76]. The search for a solution to such issues has led to the emergence of "music-oriented HCI" research, where the development of new sensing technologies, creation of mapping strategies, and user involvement in design are heavily driven by HCI know-how. Bongers, for instance, explains that the creation of sensible mappings for NMIs should be informed by research in HCI, which can help "restore the relationship between that which is felt and the sounds produced" [22]. In support of his argument, the author adapted design practices from HCI to create a theoretical framework for the design of input and output paradigms for performer-system, system-audience and performer-system-audience interactions. As another example, Salter relied in large part on existing HCI techniques to design the Schwelle responsive environment (described in Section 2.3.3 above), a project he hoped would enable him "to develop a position and framework that exploits the tensions among mapping, sonification and composition for sensor-based responsive audio environments" [164].

Some developers of NMIs took an interest in HCI research beyond mapping and interaction design, choosing instead to adopt the user-centric methodologies described previously in Section 2.4. A notable example is the work of Wanderley and Orio, who posit that "results from HCI can suggest methodologies for evaluating controllers, provided the context of interaction is well defined"[189]. Inspired by Buxton's work on the assessment of input devices, the authors argue that the user-centric evaluation of novel input devices can best be accomplished when such devices are matched to potential applications using simple, representative musical tasks. Such tasks, they add, should be designed to account for important parameters of "usability" within the context of musical performance, namely: learnability, explorability, feature controllability and timing controllability [145]. In turn, Kiefer et al. have applied Wanderley and Orio's proposed methodologies to the evaluation of various musical controllers. For instance, in a usability experiment examining the Nintendo Wiimote as a musical device, participants were asked to perform simple, musical tasks using drumming-like motions or continuous tracing gestures. Such gestures were also performed using a more established controller (the Roland HPD-15 Hand-Sonic), to provide a baseline for statistical comparisons. Participants were subsequently interviewed on their experience with both devices, as a supplement to the quantitative gestural data captured by the input devices [112]. Kiefer also conducted a similar usability experiment to assess a novel malleable controller for sonic exploration [111]. As another example, Bau et al. relied on participatory design methods from HCI for the development of the A20, a polyhedron-shaped, multi-channel audio input/output device. Throughout its design, the authors held participatory workshops where non-technical users were invited to explore the system's potential as a collaborative personal music player [14]. They were placed in groups, and asked to generate their own use case scenarios where the A20 could be used to address the theme of mobile social interaction through music. Overall, the authors were impressed by the richness of the ideas generated, many of which they had not anticipated themselves. Similarly, Geiger et al. employed participatory design techniques in the early design phase of the VRemin, a set of 3D interfaces for a virtual Theremin [76]. The findings helped the authors determine the necessary refinements required for subsequent versions of the VRemin. The Do It Yourself Smart Experience (DIYSE) Project is another instance of HCI methodologies used in the design of a musical interface. The goal of the project was to co-design novel interfaces with musical therapists that may help them in their work with people who have learning disabilities. By working closely with the therapists through the design and testing phases, the authors were able to examine the role of novel technologies in special education [132]. Most notably, however, Alexandraki and Akoumianakis based their DIAMOUSES framework, described earlier in Section 2.2.2, on a series of surveys, questionnaires, site visits and semi-structured interviews with representatives of their target end-user communities, all designed to help the authors understand the various possible contexts for network musical performance [4].

While the authors mentioned above agree that user involvement can provide the much needed structure to musical interface design, there is less of a consensus when it comes to deciding the exact nature of this involvement. For instance, while designing the VRemin, Geiger et al. found “no clear pre-existing requirements for software of this kind” and therefore had little choice but to adopt an exploratory approach [76]. The problem, in part, is that techniques borrowed from traditional HCI are applicable to the user-centric design of NMIs only to a certain degree. Kiefer et al., for instance, explain that “HCI methodology has evolved around a task-based paradigm and the stimulus-response interaction model of WIMP systems, as opposed to the richer and more complex interactions that occur between musicians and machines” [112]. In particular, the authors found that while HCI techniques allowed them to evaluate overall user performance with a controller, they were unable to capture “in the moment” data about the user experience, something they believe is important for musical evaluation. Furthermore, the practical tools examined in HCI are typically evaluated by collecting and studying quantifiable aspects of performance. However, to what extent is it possible, or even meaningful, to quantify a user’s efforts with a musical interface? As Wanderley and Orio put it, “what is the role of qualitative versus quantitative measurements in the evaluation of musical tasks?” [189]. The type of information that designers elicit from users, and the manner in which they elicit such information, have been topics of much discussion among creators of NMIs and digital arts, keen on adopting UCD methodologies. The following section sheds some light on the challenges of measuring relatively abstract concepts that are inherent to the successful design of NMIs, such as user enjoyment, creativity, expressivity and overall experience.

2.5 System Evaluation: Going Beyond Usability

According to MacDonald and Atwood, the emergence of human-computer interaction as a field in the 1960s and 1970s was, in part, driven by a growing interest in "evaluating the speed of the user rather than the speed of the system". Throughout the following decades, such an interest in user performance evolved into the notion we now refer to as "usability", commonly measured through five performance metrics: time to complete tasks, error rate, accuracy, task completion rate and satisfaction [133]. Subjects are asked to use the system under evaluation to perform a large number of short, repeatable tasks that can, in turn, be assessed quantitatively to help determine some measures of success. However, as Johnston explains, "[S]oftware designed to facilitate musical expression presents a problem in this context, as it is difficult to formulate tasks to assign to users that are measurable but also meaningful" [104]. On this matter, Cariou also notes that "it is not only undesirable but impossible to define the musician's task" [36]. Furthermore, Höök et al. argue that the "the major conflict between artistic and HCI perspectives on user interaction is that art is inherently subjective, while HCI evaluation, with a science and engineering inheritance, has traditionally strived to be objective" [96]. As a result, while the emergence of music-oriented HCI has led to a marked increase in the adoption of user-driven techniques among designers of new musical interfaces, many have found traditional usability tools to be inadequate for studying systems of an artistic nature [17, 33].

All of this has driven the need for different approaches that are better suited to the study of non-utilitarian systems. In this section, we discuss various research paradigms such as mixed-methods, the qualitative experiment and experience-based HCI, and examine a number of evaluation techniques that have emerged from the design and study of alternative types of interactive systems.

2.5.1 Mixed Research

For years, proponents of qualitative and quantitative research have stood at odds with one another, arguing for the superiority of their chosen methods. On the one hand, qualitative purists believe that social observations are as important as physical, measurable phenomena. On the other hand, advocates of quantitative research claim that such social observation cannot be objective and free of bias, thereby being detrimental by nature to their goal of establishing "time- and context-free generalizations" [139]. To further complicate matters, throughout the course of what Johnson et al. refer to as the "paradigms wars", a number of researchers have even argued that both paradigms are incompatible and should not be mixed [103, 140].

Frustrated with the inadequacies of using a single paradigm, however, a growing number of researchers began combining elements of quantitative and qualitative research in the 1980s. As this practice gained traction, methodologists began to formalize the concept of "mixed research", and examining the applicability and merits of such an approach. According to Johnson and Onwuegbuzie, mixed research is defined as "the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study"[103]. Mixed research paradigms can, in large part, be classified into two major categories: mixed-model techniques, where qualitative and quantitative approaches are mixed within or across several stages of the research, and mixed-method techniques, where an overall study includes separate qualitative and quantitative phases. Ultimately, one should aim to combine various strategies in such a way as to complement their strengths, without overlapping their weaknesses. According to Stroheimer et al., quantitative research is based on "deduction, confirmation, theory/hypothesis testing, explanation, prediction, standardized data collection, and statistical analysis", while traditional qualitative research focuses on "induction, discovery, exploration, theory/hypothesis generation" [174]. Thus, one ideal combination involves using qualitative methods to develop hypotheses that can subsequently be tested via quantitative techniques. Another approach is to conduct qualitative interviews to provide additional meaning and context to quantitative experiment data. In the end, researchers are encouraged to mix approaches individually by considering the advantages and disadvantages of each in light of the subject matter at hand. To help with such decisions, one can refer to the extensive list of strengths and weaknesses of both qualitative and quantitative research provided by Johnson and Onwuegbuzie. Table 2.1 provide a selection of items from those lists that we found to be particularly relevant to the nature of our research.

Strengths of Qualitative Research	Weaknesses of Qualitative Research
• It is useful for studying a limited number of cases in depth. • It is useful for describing complex phenomena. • Data are usually collected in naturalistic settings in qualitative research. • Qualitative researchers are responsive to changes that occur during the conduct of a study (especially during extended fieldwork) and may shift the focus of their studies as a result.	• Knowledge produced may not generalize to other people or other settings (i.e., findings may be unique to the relatively few people included in the research study). • It generally takes more time to collect the data when compared to quantitative research. • Data analysis is often time consuming. • The results are more easily influenced by the researcher's personal biases and idiosyncrasies.
Strengths of Quantitative Research	**Weaknesses of Quantitative Research**
• Data collection using some quantitative methods is relatively quick. • The research results are relatively independent of the researcher (e.g., effect size, statistical significance) • Data analysis is relatively less time consuming (using statistical software). • It is useful for studying large numbers of people.	• The researcher may miss out on phenomena occurring because of the focus on theory or hypothesis testing rather than on theory or hypothesis generation (called the confirmation bias). • Knowledge produced may be too abstract and general for direct application to specific local situations, contexts, and individuals.

Table 2.1: Selected strengths and weaknesses of qualitative and quantitative research, sampled from Johnson and Onwuegbuzie's extensive list [103].

As an example within the context of music-oriented HCI, Pras and Guastavino have successfully utilized both categories of mixed research as part of their extensive study of the interactions between musicians, record producers and sound engineers in the studio [154]. For instance, in one study exemplifying the mixed-model approach, excerpts from recording sessions were evaluated by participating musicians and producers both qualitatively (using open-ended questions) and quantitatively (using Likert scales). The goal was to determine whether the outcome of such evaluations could help musicians improve from one take to the next [153]. In another study aiming to improve communication between musicians and record producers, a mixed-method approach was used: group interviews involving both types of users were conducted before a recording session, helping them establish a common vocabulary and reach a consensus regarding the artistic direction of their collaboration. The success of these pre-production meetings was then later established using a post-production questionnaire, where participants expressed their level of satisfaction with the overall sound quality, and whether it corresponded with the wishes they expressed during the group interviews [152]. Similarly, Kiefer et al. have also relied on mixed research in their user-centric evaluation of musical controllers, described earlier in Section 2.4.3. With a philosophy influenced by HCI research, the authors conducted a number of experiments to assess performance with their controllers. Adopting a mixed-model approach, they supplemented logged quantitative data with the qualitative analysis of user comments made during and after test sessions. In the end, both types of techniques helped paint a more accurate picture of overall user performance with new musical interfaces [112].

Within the context of interactive arts, Candy et al. also encourage mixed research as a means of understanding and, in turn, improving user engagement. In agreement with our views on the complementarity of qualitative and quantitative methods, the authors state that while quantitative methods can help verify a hypothesis, qualitative methods are useful for developing hypotheses and detailed insight on specific cases. Therefore, while studying user experience with *beta_space*, an experimental environment at the Powerhouse Museum in Sydney where the public can engage with the latest research in art and technology, the authors combined three types of data collection methods: direct observations to investigate user behaviours, and questionnaires and interviews to investigate intentions and reflections behind those behaviours. While the results of the questionnaires could be quantified directly, careful qualitative analysis was applied to the observations and interviews. First, the observations were acquired by means of a context analysis, a technique whereby researchers present at the scene record, in as unobtrusive a manner as possible, events that take place in field diaries supplemented by audio recordings. Subsequently, the data produced through such diaries and recordings, along with the answers obtained separately through the user interviews, were examined via the content analysis technique [33]. Content analysis, a popular methodology in the social sciences, allows researchers to derive information from non-numeric data. At its core, content analysis operates on the principle of grounded theory, or the notion that hypotheses are contained within and can be induced from data collected during an experiment. This is in contrast with traditional (and typically quantitative) scientific research, which postulates that hypotheses should be clearly formed before an experiment. While interpretation of verbal and behavioural data is subjective by nature, content analysis introduces a certain level of rigour to the process: the process relies heavily on a procedure known as coding, during which codes, or tags with pre-defined meanings, are assigned to events in a data set (such as behaviours obtained from user observations, or quotes obtained from user interviews). Coding is typically applied in an iterative fashion, whereby codes deemed similar enough are grouped and combined, until a smaller, relatively stable set of codes emerges. From this resulting set of codes, researchers can subsequently begin to draw some of the ideas behind the broader user motivations, tendencies and goals.

Mixed techniques provide an "expansive and creative" approach that not only helps researchers overcome the inherent limitations of individual methodologies, but also promotes collaboration across multiple disciplines. In fact, a growing number of HCI researchers are mining the social sciences for techniques that can help supplement the quantitative approaches for which the field is renowned. Nielsen, for instance, states that "[i]t's a dangerous mistake to believe that statistical research is somehow more scientific or credible than insight-based observational research." Explaining that a fixation on numbers can lead usability studies astray, the author adds that qualitative studies are "less likely to break under the strain of a few methodological weaknesses" [141]. One technique in particular that has proven to be quite effective due to its strong procedural nature is the qualitative experiment.

Its adoption by the HCI community is discussed in the following section.

2.5.2 The Qualitative Experiment

Despite its long history in the social and natural sciences, the term "qualitative experiment" was not formally defined until the 1980s, largely through Kleining's analysis of existing scientific methods. According to Kleining, a qualitative experiment is "[t]he intervention with relation to a (social) subject which is executed following scientific rules and towards the exploration of the subject's structure. It is the explorative, heuristic form of an experiment" [116] (translated by [155]) . The qualitative experiment begins with theorizing the existence of relationships and processes that are difficult to quantify, and that can only be quantified after additional special treatment. Subsequently, variables deemed related to such relationships and processes are examined in rigorous experimental settings analogous to those used in quantitative experiments. However, where the qualitative experiment differs from its quantitative counterpart is in the nature of the data collected: interviews, discussions, case studies and diaries are some examples of the techniques commonly used to elicit user feedback in the qualitative experiment. Ideally, the results of the qualitative experiment should serve to develop hypotheses that can, in turn, be verified through quantitative studies. Thus, both approaches can effectively complement each other, providing, as mixed research typically does, a more complete picture of the subject matter under consideration.

The exploratory nature of the qualitative experiment renders it quite suitable for developing mental models of user interaction with systems that are completely novel, or that employ new technologies that have yet to be fully understood or documented. In addition, the qualitative experiment can be useful in understanding hedonic qualities of interaction that may be difficult to quantify. As a result, qualitative experiments have enjoyed a growing popularity with the music-oriented HCI community for a number of years. For instance, Johnston et al. conducted a qualitative experiment to test novel software musicians could use while playing traditional acoustic instruments to create a mix of computer-generated and acoustic sounds, as well as associated visuals. The authors felt strongly about their choice of study, explaining, "If the aim had been to produce a general-purpose musical instrument for performing music in a well-established tradition, then this would be simpler. Tasks such as playing a scale, trilling etc. could be assigned and measurements to ascertain how successfully users are able to execute them." However, the software was specifically designed to "disrupt habitual ways of thinking about music" and, in turn, encourage musicians to explore new ways of playing and composing [104]. Thus, a quantitative study could not have adequately helped evaluate the defining characteristics of the virtual instrument. The authors therefore designed a qualitative study anchored on content analysis, and encouraged the seven professional musicians who interacted with their system to "think out loud". All sessions were filmed. Furthermore, an observer attended each session to take notes and provide additional perspective. Subsequently, the authors utilized grounded theory techniques to explore a two-fold question: how the musicians approached the virtual instrument, and how it affected the music-making process. The procedure Johnston et al. used consisted of four steps: transcribing all video footage (including non-verbal incidents); line-by-line open coding of the transcripts; memoing, or noting relationships as they emerge during coding; and sorting, or organizing the memos to identify core issues. Overall, Johnston et al. found that their process did not necessarily help them understand how to design better instruments. Nonetheless, it was very useful in determining what musicians did and did not like about the virtual instrument. Most notably, however, the authors maintain that "[e]ngaging in loosely structured dialog with expert creative users is effective in building an understanding of the sometime complex ways in which they interact with software while engaged in creative work" [104].

Designers of interactive installations are also fast becoming proponents of qualitative user evaluations. For instance, throughout their study of user interaction and engagement with interactive arts, Bilda et al. have noted that experience with an artwork can be as unique as the work itself and, as a result, found that the appropriate evaluation technique must be tailored to match each system under examination. As a result, during the design of GEO Landscape, an installation that allows users to create a story by moving through a digital landscape, the authors relied on qualitative evaluations of user experience to improve their system design. Their approach was based on a desire to understand the emotional aspects of interaction, which the authors felt they could not adequately capture using the "too regimented" existing usability techniques [16]. Therefore, they designed a two-stage contextual inquiry to examine the interactions of six novice and six expert users with the system. During the first stage of the inquiry, participants were invited to explore the system freely. Following this initial introduction, they were asked to show or explain how they interacted with the work. This segment helped determine whether users did in fact understand the various components of the interface. The second stage of the inquiry encouraged participants to explore the system again, this time while expressing any opinions about newly discovered aspects of the work. Throughout that stage, the researchers also directed additional questions to the participants to determine whether they fully uncovered all the layers of the work, and considered their meanings. Both stages were filmed and a content analysis was performed to outline the similarities and differences between the novices' and experts' reactions and preferences. The findings helped the authors understand how to refine GEO Landscapes in such a way as to better sustain user interactivity and, in turn, increase audience engagement with the work.

Developers of new musical interfaces and interactive installations are not the only ones to embrace qualitative evaluation as an essential component of the user-centric doctrine: despite the ubiquity of quantitative experiments in HCI, a number of researchers have recently begun advocating the unique advantages that qualitative methods stand to offer designers of practical systems and applications. Most notably, to promote the use of rigorous and procedural qualitative methods, Ravasio et al. created a formal framework for the qualitative experiment specifically tailored for HCI researchers. The authors provide six possible strategies for observing test-dependent variables in a qualitative setting [155]:

- Separation/Segmentation: The system is partitioned into various sub-parts, and each part is then isolated from the whole and examined.

- Combination: The system is combined in a new way with another system. One then examines the extent to which the systems are different from or compatible with one another.
- Reduction/Attenuation: Stepwise, individual functionalities of the system are removed or attenuated. One then studies the effect this has on the overall system.
- Adjection/Intensification: The reverse of reduction/attenuation, where aspects of the system are intensified and their impact examined.
- Substitution: Certain parts of the system are replaced by new ones. One then studies instances where a small substitution has a large impact, or a large substitution has a small impact.
- Transformation: The whole system is transformed, with only a handful of old attributes remaining the same. The impact of such change is then examined.

Ravasio et al. advocate using the qualitative experiment when one's goal is "to discover (rather than to verify) structures, procedures, processes and their interdependencies, and when the setting should be as close as possible to real-life ... but still require a degree of controlled removal of context" [155]. Together, the strategies they propose offer systematic means of uncovering relationships between the parts of a system, and can help in determining which aspects of the user experience each of these parts is most likely to impact. In addition, due to its exploratory nature, the qualitative experiment lends itself quite naturally to the study of those rather subjective facets of interaction that constitute a large part of the user experience, as discussed in the following section.

2.5.3 Experience-Based Design

When Gould and Lewis introduced their "key principles" for usability design, they advocated for an early focus on user tasks. However, with a growing number of varied disciplines turning to HCI research for guidance on designing not only usable, but also engaging systems, many researchers were faced with the shortcomings of this task-based approach. MacDonald and Atwood, for instance, argue that "[a]s use contexts have broadened and technologies have become more pervasive, designers and evaluators recognized the importance of considering the "non-utilitarian" aspects of using computers, which shifted the focus from task-based performance to user affect and the value of computer interaction in everyday life" [133]. Furthermore, with the introduction of the human-centered perspective, Bannon adds that technical and functional characteristics had become insufficient, and calls for a better framework for conceptualizing human activities both at the interpersonal and behavioural levels [12]. Kaye et al. further posit, "what of technology not for accomplishing tasks but for having experiences, for expressing one's identity, for flirting and arguing and living?"[110]. The authors add that evaluating solely on usability is "to miss the very point of these technologies". Such frustrations, in turn, led to the emergence of what is now known as "third-wave" or "third-paradigm" HCI, a trend described by Kiefer et al. as a "a response to the evolving ways in which technology is utilised as computing becomes more increasingly embedded in daily life" [112]. Third-wave HCI promotes an experience- rather than task-based approach to user-driven design. It encourages what Fallman and Waterworth describe as a focus on "experiences rather than performance; fun and playability rather than error rate; and sociability and affective qualities rather than learnability" [66]. As a result, third-wave HCI is particularly suited to the design and evaluation of novel interactive musical interfaces. This view is supported by Blaine and Fels, for instance, who advise designers of collaborative musical interfaces to replace traditional music metrics based on melody with ones that place more emphasis the musicians' overall experience [19]. Within the context of interactive arts and installations, Bilda et al. add that studying engagement is "not just about fulfilling a goal or a series of tasks; rather it is more about what a participant feels and experiences" [16].

In addition, experience-based HCI has proven quite effective in the design and evaluation of embodied interfaces. For instance, throughout the design of the Ariel system (described earlier in Section 2.3.4), Corness and Schiphorst concentrated on "intuition, empathy and intention as key elements in interaction" [55]. As such, their evaluations of the system consisted of long-term user studies where, rather than complete specific tasks, performers were asked to improvise with the system for multiple sessions, and explore such issues as trust, presence, connection and communication. Similarly, in their exploration of movement awareness in ubiquitous computing, Levisohn and Schiphorst advocate an emphasis on learning, enjoyment and aesthetics as crucial to the successful design of embodied interaction [124]. As another example, Loke and Robertson describe the design of embodied interactions as one with a focus on the "the movement, bodily awareness and felt experience, which account for such "ineffable" qualities of human experience; qualities which can often escape definition or measure, but are a necessary part of meaningful experience" [130].

The evaluation of the user experience, however, remains somewhat of a challenge. Fallman and Waterworth, for instance, argue that it is necessary to "to find or develop appropriate and mature procedures for gathering and analyzing empirical data in relation to these new, experience and meaning-related aspects of interacting with computers" [66]. However, as MacDonald and Atwood explain, evaluators face a "lack of shared conceptual framework" for the user experience. The authors argue that it is common practice to associate non-instrumental or hedonic goals with experience, and instrumental or pragmatic goals with usability. As a result, the evaluation of experience has typically focused on hedonic attributes, while the evaluation of usability attempts to capture pragmatic ones. Nonetheless, some researchers have found that the user experience encompasses both the hedonic and pragmatic aspects of system use. Kaye et al., for instance, explain that while the evaluation techniques for task-focused measures, such as "classical notions of usability", are inadequate for the evaluation of experiences, they are still far from unnecessary [110]. In fact, O'Brien and Toms have identified perceived usability, aesthetics, focused attention, felt involvement, novelty, and endurability to be six attributes that constitute engaging experiences according to existing literature on the topic [144]. In addition, Hassenzahl and Ulrich found that the inclusion of active instrumental goals in system evaluations had an impact on the way users perceived their overall experience [90]. As a result, both the pragmatic and hedonic facets of interaction should ideally be examined during experience evaluation. While pragmatic goals can typically be quantified according to such metrics as task completion time, accuracy and error rates, the study of hedonic qualities, such as “fun, pleasure, goodness and beauty” [17], can be more difficult. Such challenges are explored in the following sections.

Affective Computing

Typically, musicians and artists express a greater interest in the hedonic aspects of their experience with a system than they do for the system’s efficiency or practicality. However, a number of researchers have demonstrated that users of practical applications also exhibit a strong appreciation for other, less pragmatic qualities of interaction. One particular area of HCI notable for reaching beyond the traditional notion of usability is that of affective computing. A relatively modern field originating from Picard’s seminal 1995 paper, affective computing focuses on the development of tools that can recognize, process and, in some cases, even simulate human feelings and emotions. As Norman explains, emotions can change the way we approach a problem, making cognition and affect, processes that lead to understanding and evaluation respectively, a “powerful team” [142]. As a result, a growing number of HCI practitioners are exploring what Chateau and Marisol describe as “the emotional dimension of Computer-Human Interactions”, in an attempt to unlock its potential for vastly improving user interaction with a system [45].

Due to its ability to “change our emotional state”, aesthetic appeal is one characteristic in particular that has proven to be quite critical to the study of affective systems [142]. In fact, Lavie and Tractinsky suggest that visual aesthetics are a strong determinant of users’ satisfaction and pleasure with a system [120]. As an example of this notion, Tractinsky et al. uncovered a strong correlation between users’ perception of the aesthetics of an interface and the usability of an entire system, which was, in their case, an automated teller machine [186]. Similarly, Hassenzahl et al. found that novelty and originality played a substantial role in the overall appeal of a GUI they designed to control a pump in an assumed industry plant [91]. This idea of “aesthetic interaction” has in fact been examined extensively by Petersen et al., who explain that "it is not about conveying meaning and direction through uniform models; it is about triggering imagination, it is thought-provoking and encourages people to think differently about the encountered interactive systems" [149]. Putting that notion to practice, the authors conceived of the WorkSPACE project, a system that envisions a pervasive computing environment using walls, tables and floors as interactive surfaces for the exchange and manipulation of documents. Users interact with the system using a ball, an artifact that implies playfulness and a forgiving attitude towards erring. They are encouraged to explore the system, appropriate it, make mistakes, establish their own relationships with digital materials and, most importantly, enjoy themselves throughout.

While the importance of affect in the design of engaging system is widely acknowledged, there is less of a consensus on how such a quality is best evaluated. Isbister, for instance, explains that "[e]valuation of user affect is a domain that is not as well articulated and explored as is assessing whether a system is usable, or whether it actively increases work productivity" [98]. To this, Hassenzahl et al. add that "[t]raditional usability engineering methods are not adequate for analyzing and evaluating hedonic quality and its complex interplay with usability and utility" [89]. As a result, a number of researchers have designed novel evaluation techniques that specifically introduce a greater a level of rigour to the study of the more subjective aspects of interaction, a number of which are listed in Table 2.2 below. We note that while the techniques listed here could potentially be applied within a musical context, Kiefer warns that they first "need to be assessed specifically in terms of evaluation of musical experience as well as user experience" [112].

Name	Motivation	Description
Repertory Grid Technique (RGT) [66]	To empirically elicit and evaluate a user's subjective experience when interacting with technology.	RGT is a matrix whose rows contain qualitative constructs, and columns represent elements under investigation. What makes RGT unique is that it allows subjects to create their own personal constructs, or bipolar dimensions, according to what they deem most important when it comes to the various elements being investigated.
Semantic Differential[89]	To reduce the efforts required in creating personal constructs as part of the RGT technique.	Like RGT, the semantic differential is a matrix of qualitative constructs and elements under investigation. Where it differs, however, is in its reliance on seven pre-selected constructs that best characterize hedonic quality for all subjects.
AttrakDiff[92]	To provide a tool for large-scale studies utilizing semantic differentials to examine the pragmatic and hedonic qualities of interactive systems or products.	AttrakDiff is a free web-based tool that offers experimenters the possibility of organizing large-scale comparative studies of interactive systems. It allows for the comparison of entirely different systems (A/B), or of different iterations of the same system (before/after). After testing each interface, subjects are asked to rate their experience using a semantic differential. Subsequently, AttrakDiff generates for the experimenter graphical representations of the aggregated user perceptions of pragmatic and hedonic qualities of the system(s) under investigation.
Structured hierarchical interviewing for requirement analysis (Shira)[89]	To explore the meanings behind a pool of attributes commonly used to describe emotional response to a system.	Each participant is introduced to a system under evaluation, then asked to select an attribute from a pool that includes common terms such as "controllable", "innovative" or "simple". Subsequently, she is asked to list any software features that warrant attaching that attribute. This represents the context level. In the following step, the participant must provide recommendations for improving each entry in the context level. This represents the design level. Together, the attributes and levels result in a hierarchical personal model for each participant that, in turn, can be of use during a system's early design stage.

Sensual Evaluation Instrument (SEI) [98]	To gauge emotional response in a non-verbal manner, thereby preserving the rich and multi-layered nature of the feelings experienced by users interacting with a novel system.	SEI consists of a set of biomorphic, sculpted objects whose contours express affective qualities. The forms of the objects were designed to be familiar on a visceral level, yet still open for interpretation. Users are asked to select the objects that best represents their feelings while interacting with the system under examination. Their choices are subsequently discussed in a post-test interview, and later analyzed by the experimenter.
Product Emotion Measure (PrEmo) [57]	To measure momentary reactions to a product, in as fast and intuitive a manner as possible.	PrEmo is a self-report instrument that utilizes 18 different animations of the same cartoon character, each expressing a different emotion. Subjects are asked to select one or more animations that best reflect their instantaneous emotional response as they are shown pictures of the product(s) under evaluation.
AMUSE [45]	To create a complete picture of a user's emotional reaction to a system through the synchronization of multiple data collection methods.	AMUSE is a computer tool that records signals from eight electro-physiological sensors, an eye-tracker, a mouse and keyboard tracker, windows displayed on the computer, and video of the user. It also allows the experiment to mark events of interest during a test session. Subsequently, the experimenter can sync all data streams on one screen, and analyze them simultaneously to gain a full picture of the user's emotional response.

Table 2.2: Overview of various techniques designed specifically for the evaluation of affect.

Fun, Pleasure and Flow

Fels describes a "well-designed instrument" as one comprising an interface that is constrained and simple enough to allow a novice to make sounds easily, while also remaining challenging enough for the experienced player to explore a path to virtuosity [68]. A similar view is also echoed by Wrigley and Emmerson, who argue that "musical activities must provide players and composers with continually demanding challenges in such a way as to keep the individual interested, stimulated and in flow" [193]. This notion of "flow" was first formalized by Csíkszentmihályi to denote a state of optimal experience marked by a feeling of energized focus, full immersion and enjoyment. [89]. A state of flow occurs when a task presents a level of challenge that is perfectly matched to a user's skill set, thereby precluding overstimulation on one end, or boredom on the other. It is typically characterized by a sense of reward, a merging of action and consciousness and, naturally, a notable level of pleasure. In fact, Csíkszentmihályi originially conceived of flow while investigating the concept of enjoyment itself. As MacDonald et al. explain, "[i]n this flow state, people experience a narrow field of intense concentration, they forget about personal problems, feel competent and in control, experience a sense of harmony and union with their surroundings, and lose their ordinary sense of time" [134]. Thus, flow can be considered a reliable indicator of pleasure and enjoyment.

Another quality that a number of designers of interactive systems have turned to recently is the notion of "fun". As Bushnell notes, "[w]himsy and fun are often the precursors to powerful tools that are used later for more serious applications" [29]. Furthermore, Hassenzahl et al. consider "joy of use" to be an important dimension of overall usability that designers must consider, if only for the humanistic view that "enjoyment is fundamental to life" [89]. The authors add, however, that "[t]here is an explicit difference between knowing that hedonic quality could play a role in designing interactive systems, and actively accounting for it". This could perhaps be attributed to the fact that such notions as "pleasure" or "joy" are too nebulous to define accurately. Furthermore, Blaine and Fels feel that while pleasure is clearly observable, assessing the metrics of fun can be more ambiguous [18]. Too this, Wiberg adds that "we have so little knowledge about how traditional usability evaluation works in the context of fun and entertainment work, it is difficult to argue for new approaches" [192].

While such challenges receive relatively limited attention in HCI, they have been widely examined within the study of gaming, a field otherwise known as ludology, where the investigations of fun, pleasure and flow are considered to be a cornerstone of game design. LeBlanc, for example, has proposed "Eight kinds of fun", a taxonomy that identifies various sources of pleasure, such as sensation, challenge or discovery, that video game players typically experience [121]. As another example, Sweetser and

Wyeth capitalized on the notion of flow when they created the “GameFlow” model, in a bid to better “design, evaluate and understand enjoyment in games” [176]. Combining common heuristics from the gaming literature, GameFlow is composed of eight elements: concentration, challenge, skills, control, clear goals, feedback, immersion and social interaction, each of which comprises criteria deemed necessary towards ensuring that players can reach a state of flow. To quantify the assessment of flow even further, IJsselsteijn et al. developed the gaming experience questionnaire (GEQ), a set of in-game and post-game questionnaires targeted towards evaluating competence, sensory and imaginative immersion, tension, challenge, and negative and positive affect, all of which can be indicators of enjoyment levels. The GEQ also includes a “social presence” questionnaire to measure the empathy, negative feelings and behavioural involvement amongst players [97]. Most of the facets of flow and pleasure that the techniques listed here examine are, in fact, commonly experienced across a wide variety of activities. As a result, we argue that these methods may be extended beyond the field of ludology and, with some modification, used effectively to evaluate and improve the user experience with many types of interactive systems, including new musical interfaces.

Creativity

Creativity has always been considered an essential quality of most, if not all, artistic endeavours, including musical performance. In recent years, however, creative engagement has come to be regarded as an important quality to consider when designing interfaces meant not only for artistic purposes, but utilitarian ones as well. For instance, Candy notes that there is a growing demand for information technology tools that can better support the needs of “creative users”, such as professional knowledge workers. These users are increasingly relying on computers to facilitate creative aspects of their work, and in order to meet their demands, Candy explains that “an understanding of the nature of creative cognition as well as an evaluation of the tools that are used in the creative process is needed” [34]. Similarly, Vass et al. stress the importance of supporting creative engagement in problem solving environments (PSEs). Interestingly, the authors state that "[t]he difference between a problem solving user and a creative problem solving user is the presence of flow", a characteristic that, as discussed in the previous section, can also be an indicator of pleasure [187]. As a result, they designed Workflow, a model that incorporates the principles of flow into the problem solving process. By extending the concept of usability to include creativity, Vass et al. intended for their model to improve the design of traditional PSEs, and ultimately enhance the overall user experience with such tools. As another example, Terry et al. examined the effect of various interface designs on the creative process during image manipulation. They found that creativity is best supported when the user is able "to experiment; to explore variations; and to evaluate past, current, and potential future states" of the set of images at hand [184]. A similar sentiment was echoed by Luhtala et al., who in explaining the motivation behind the DIYSE project (described earlier in Section 2.4.3), express an increasing demand for "new design tools that enable creativity by means of explorative interaction, as opposed to limited executive and mission based interaction" [132].

The evaluation of creativity, however, continues to be an open problem. As Candy explains, "a creative act is by its very nature, neither predictable nor repeatable" [34]. To this, Kiefer adds that "getting people to perform a precise task can be difficult especially when you have creative people performing a creative task"[112]. Within the context of interactive art, Bilda and Edmonds expand on this idea, stating that "[b]y its very nature, creative engagement with interactive art systems is as varied as the individual people who interact with it ... and, therefore, quite difficult to predict" [17]. Therefore, while these authors acknowledge that a user-centric approach can be highly beneficial to the design, evaluation and improvement of systems that promote creative engagement, they also recommend approaching the selection or design of any evaluation methodologies with great care. In fact, throughout their work on the *beta_space* project, Candy et al. opted for a "practice-led" approach, where a studio environment was re-created in a research setting, allowing the authors to examine artists, curators, audience members and even researchers themselves involved in one aspect or other of the creative process. Various techniques such as the think-aloud method, where subjects are encourage to verbalize their thought process, and the co-discovery method, where subjects discuss their interactions naturally with one another, were combined with traditional interviews, questionnaires and observations to give researchers a thorough understanding of the various facets of creative engagement.

2.6 Synthesis of Literature Review

While it was once seen as a mere channel for the exchange of digital information, a growing number of artists and researchers are embracing the network, with all its idiosyncrasies, as a unique performance environment. Nonetheless, like many on-line collaborations, distributed performance exhibits a decreased level of social interaction when compared to its co-present counterpart. Kapur et al., for instance, lament "the loss of the identity of the 'band' itself, that is, the interaction of a finite number of players, each with their unique role, playing together on a single stage" [106]. In that sense, distributed performance shares a common challenge with CSCW, an application area that also often exhibits a decreased sense of mutual awareness and spontaneous interaction [61, 171]. Furthermore, traditional research on distributed musical collaboration has often focused on latency, with early works investigating solutions for reducing latency, and later ones exploring the merits of accepting it. We, however, argue that resolving issues of latency can only help musicians overcome their remoteness to a limited extent: successful collaboration over a network depends not only on a system's underlying technology, but also the degree to which its interface can support and even augment existing interactions in a meaningful way, a philosophy to which Ackerman refers, within the context of CSCW, as bridging the "social-technical" gap [1]. Finally, in the most ideal of scenarios, a distributed performance system should help displaced musicians feel as though they are all present within the network, interacting with one another in a shared space. As a result, we posit that shared workspaces, as seen in the TeamWorkStation [99] and ClearBoard projects [101] described in Section 2.1, exemplify the philosophy of being "in" the network found in the literature on distributed performance, and detailed in Section 2.2.1. The ideas described here lead us to regard distributed performance as a unique application area of the "same time/different place" category of CSCW [102]. Therefore, as has been the proven case within the context of CSCW, supporting distributed musical collaboration in a manner that engenders a sense of co-presence should begin with a thorough understanding of the ways in which participants interact and communicate with one another. Since positioning a target user at the centre of one's design efforts is a practice emblematic of HCI research, we turn to this field for further insight.

Tanaka explains that the design of new musical interfaces "should benefit from techniques from human-computer interaction research" [181]. Many researchers in the field of music technology would agree, as is evident from the substantial body of "music-oriented HCI" work discussed in Section 2.4.3. Traditionally, much research in this area was devoted to using knowledge from HCI to match input/output paradigms suitably to musical tasks. However, an increasing number of NMI designers are turning to user-centric techniques, another fundamental area of HCI, as a means of refining their work. Nonetheless, there is an apparent lack of established conventions when it comes to conducting systematic evaluations of NMIs. As Wanderley and Orio point out, the wealth of creativity seen in the design of novel controllers, environments and interfaces is countered by "the lack of commonly accepted methodologies for evaluating existing developments" [189]. Similarly, Poepel describes the methods for successfully evaluating the usability of NMIs as being "rare" [150]. This problem lies in large part in the objective, quantifiable nature of performance indicators traditionally examined in HCI task-based system evaluations. In contrast, much of a musician's experience with an NMI can often be dictated by qualities that are subjective in nature: pleasure, creativity, aesthetic enjoyment and engagement, all of which we argue cannot be quantified directly. As a result, some music technology researchers began adopting methodologies from the social sciences and designing qualitative experiments to better understand these subjective aspects of performance. Such a practice has also for some time been common among designers of interactive installations who are keen on improving and sustaining their audience's engagement through user-driven techniques. In fact, we are witnessing a shift away from traditional, task-based, usability-driven design, and towards third-wave

HCI, which promotes experience-based design and evaluation, particularly within creative and artistic contexts. As described in Section 2.5.3, the experience-based approach has become increasingly common among designers of interactive arts, musical interfaces and playful systems keen on adopting what Bannon describes as a "human-centered perspective" [12]. Such researchers are utilizing quantitative and qualitative techniques, and even developing entirely new tools, in an effort to create experiences that can closely match their target users' needs and expectations.

Orio et al. state that a "bi-directional flow of knowledge between classical HCI research ... and the design of new computer-based musical instruments can lead to improvements in both fields" [145]. However, while the field of music technology has long benefited from research in HCI, the export of knowledge from music technology back to HCI has comparatively, to the best of our knowledge, been explored minimally. The most notable example is perhaps Buxton's treatment of interaction paradigms through their constituent, low-level elements. Buxton has advocated for using "tension and closure to develop phrase structure to our human-computer dialogues", encouraging us, as an example, to view the components of the move, or cut and paste, command as "being woven together by a thread of continuity similar to that [sic] binds together a musical phrase" [30]. This compelling example encourages us to further mine music and performance for know-how that may benefit interaction design. In particular, we expect that studying interaction and exploring interface design within creative or artistic contexts could help us learn how to improve the user experience within a wide variety of applications, a view supported by a growing number of researchers. For instance, in explaining some of their motivation behind the Ariel musical system described earlier in Section 2.3.4, Corness and Schiphorst argue that the knowledge developed from researching a performer's embodied knowledge could be transferred to other domains of digital interaction [55]. Similarly, Loke and Robertson posit that their study of dancers' movements could "offer possibilities for opening up the landscape of the experiential, moving body in the design of new forms of movement-based interactive technologies" [130]. As another example, Höök et al. believe that "the perspective of artists can help HCI evaluation by suggesting some new aspects of the relationships between system builders, users, and evaluators" [96].

Bongers also explains that "[t]he general field of HCI research can benefit from findings in the field of electronic art, where the inherently unorthodox approach leads to experiments involving a rich and intimate interaction between people and electronic systems" [21]. Finally, Tanaka makes an analogous observation within the context of music, stating that "[i]nteraction patterns observed in music could in fact inform technology design. Music is a cultural practice that has the potential ultimately to contribute to a deeper understanding of interaction" [181]. We strongly echo these sentiments and, in turn, wish to explore new and meaningful interaction paradigms through our user-driven approach to the design of a novel environment for distributed performance.

Informed by the various fields surveyed in this literature review, we sought to develop a responsive environment that could leverage existing musician interactions towards improving a unique form of distributed collaboration such as network musical performance. Furthermore, inspired by the parallel ideas of the network as a "space for being" [167] and collaboration within "shared workspaces" [99], we wanted our system to offer participants the illusion of performing within a "shared space", in a bid to confer a greater level of co-presence than is possible with standard distributed performance solutions. Finally, we strove to design such a system entirely from a user-driven, experience-based perspective that builds on the qualitative assessment techniques defined by Ravasio et al. [155], an understanding from Csíkszentmihályi of the importance of supporting flow [56] in activities such as musical performance, and an overall emphasis on creative engagement.

Chapter 3

Preliminary Work and Motivation

3.1 The Videoconferencing Privacy System

The Videoconferencing Privacy System (VPS) was a project that would become an inspiration and precursor for our responsive environment for distributed performance. It was designed to resolve the challenge of supporting privacy in a videoconference setting through gesture detection. Turning one's head towards someone while cupping a hand around one's mouth is usually understood to convey a desire to speak privately with this person. Our definition of "privacy", within the context of online collaborative environments, was adopted from Fernando et al., and refers both to avoidance of divulging confidential information or secrets, as well as to the freedom from disturbance by interruptions or discussions that one deems as irrelevant to oneself [72]. Privacy is important not only for social reasons, but also for attention filtering, such as attending to a more important source of information. To this end, the VPS used low-cost, widely available hardware to detect the semiotic gesture described above, and responded by establishing private sidebar communication between two videoconference participants.

3.2 Design and Implementation

3.2.1 Scene Rendering

The graphical environment for the Videoconferencing Privacy System was rendered using the Shared Reality Lab's Audioscape platform,[1] which utilizes elements of OpenSceneGraph[2] and Pure Data [3] to manage complex 3D audio scenes. To transport the video between endpoints, we employed McGill's Ultravideoconferencing software,[4] with an added module to perform background removal for effective blending of the remote participants into the local user's virtual scene, as illustrated in Figure 3.1. Background removal allows the projected video image of each remote participant to seem as though they are physically part of the environment seen by the local user, thereby increasing the overall sense of co-presence. The audio, on the other hand, was transmitted amongst participants through Pure Data's `~nstream` object,[5] which allows Pure Data to share multichannel uncompressed audio amongst machines through UDP at a low latency.

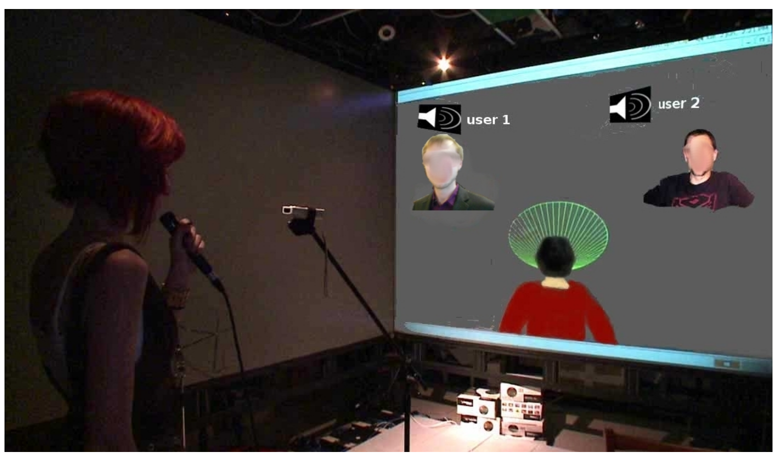

Fig. 3.1: The Videoconferencing Privacy System's graphical environment

[1] `http://www.audioscape.org/`
[2] `http://www.openscenegraph.org/`
[3] `http://www.puredata.info/`
[4] `http://ultravideo.mcgill.edu/`
[5] `http://cim.mcgill.ca/~nicolas/downloads.html`

For feedback purposes, each user was rendered locally as an egocentric avatar with a projected vocal cone from the mouth, indicating the scope of audio propagation to the other participants (see Figure 3.2). Additionally, the avatar's head direction followed that of the user.

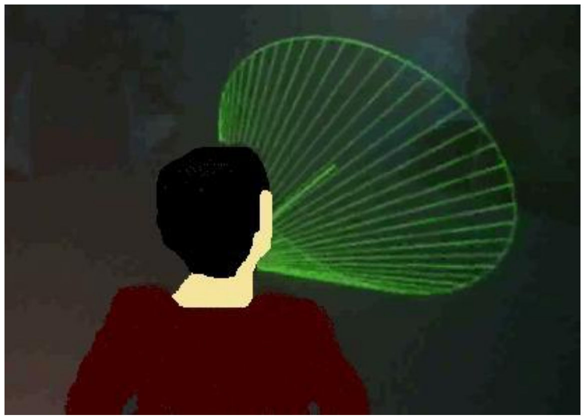

Fig. 3.2: Local representation of each user as an egocentric avatar with directional vocal cone.

3.2.2 Gesture Detection

As we wanted to ensure that the Videoconferencing Privacy System could be deployed easily at any number of locations, we opted to use the Nintendo *Wii Remote*, or Wiimote, rather than expensive motion capture systems, to track and interpret gestures. The Wiimote includes an infrared (IR) sensitive camera that can report, over a Bluetooth channel, the relative pixel coordinates along the x- and z-axes of the four brightest IR light sources in its field of view. Participants were asked to wear a custom-made "sensor strap" around their wrist, as pictured in Figure 3.3. The strap contains a set of IR light-emitting diodes. A Wiimote was fixed in place to detect the lights on the strap. Subsequently, the z-axis position values allow us to determine whether the hand is raised in a cupping gesture (see Figure 3.4), while the x-axis values reveal the approximate direction the user is facing. The latter is then used to resolve which remote participant's 2D video representation the local user is turned towards, which in turn indicates the person with whom she wishes to hold a sidebar conversation. Once this information is collected, the audio transmitted by the user is only received by the participant whose video representation she is facing. All other participants cannot hear what is being said by the user until she lowers her hand, thereby signalling, for the time being, the end of her need for privacy.

Fig. 3.3: The custom-built sensor strap built for the Videoconferencing Privacy System.

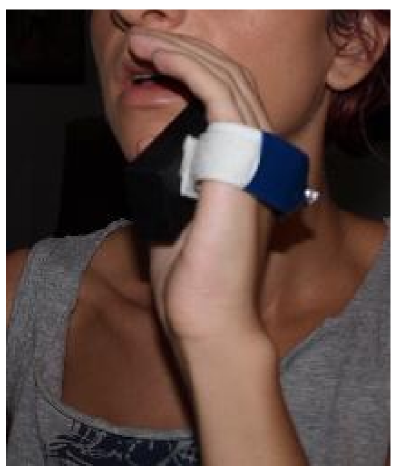

Fig. 3.4: A user wearing the sensor strap, with her hand cupped around her mouth.

3.3 From Videoconferencing to Network Musical Performance

The Videoconferencing Privacy System was a prototype for a responsive environment that, through a simple and inexpensive hardware configuration, offered its users a higher level of control than traditional videoconferencing systems. By taking advantage of commonly understood social gestures, we were able to augment videoconferencing with an interesting functionality that required minimal learning and setup. As we were implementing the VPS, however, we noted that individuals holding sidebar discussions within a group are somewhat reminiscent of musicians “jamming” interdependently within an ensemble. Therefore, we began to contemplate whether it might be possible augment distributed performance in a comparable manner, and capitalize on common interactions between musicians to offer them more responsive and creative performance environments.

We quickly realized, however, that moving our prospects from the field of traditional videoconferencing to that of network musical performance would prove challenging. Semiotic gestures are not as clearly defined and understood in musical performance as they are within the context of social situations. In addition, while Wiimotes and sensor bars proved to be suitable for a videoconferencing setting, they could only be used effectively within a limited range. Musicians, in contrast, expect to be able to move around rather freely in order to perform in an expressive manner. Finally, while notions of privacy and selective inclusion are integral to most social interactions, their direct equivalents within the context of performance are less evident. As is illustrated by the wide scope of interfaces discussed earlier in Chapter 2, the choice of controls afforded by novel musical systems can often be a simple matter of personal preference. Thus, to what extent could we successfully determine *a priori* the types of interactions musicians would find desirable within a distributed context?

In order to explore these challenges and their implications, we decided to undertake the design of a novel system for augmenting distributed performance from a user-driven perspective. To narrow down the scope of our task, we began with an outline of high-level goals we believed our system should ideally meet. Namely, these goals are to:

- restore the social aspects of performance, which are too often lost in a distributed setting, by increasing the level of interactions among participants
- provide a platform for exploring new interaction paradigms in the distributed context
- offer musicians novel, dynamic controls
- avoid creating a performance scenario that simply mimics the co-present context
- promote network musical performance as an alluring and unique concept in its own right

Furthermore, to better guide our design of all functions and controls while effectively meeting the goals listed above, we determined that all system features should be:

- designed with a focus on the musicians by means of user-driven techniques
- driven by embodied performer-performer interactions
- controllable in a dynamic and seamless manner that does not necessitate the user to detach himself from the higher-level task of performance
- transparent and, whenever possible, utilizes mappings that adhere to a clear and common metaphor
- easy to learn and remember.

In sum, our approach strives to incorporate the “walk up and play” philosophy of Interactive Installations, the collaborative nature of Interconnected Musical Networks, and the fluid nature of responsive environments into the context of distributed performance. However, we did not want our contribution to be simply yet another novel musical controller. As discussed earlier, music is a particularly challenging application area for HCI. The nature of performance forces us to re-evaluate our definitions of user goals and tasks, and calls for non-traditional input and output paradigms. In fact, as system designers, we stand to learn quite a bit from observing and working closely with such a unique user as the musician. Therefore, we hoped that, ultimately, lessons could be drawn from our efforts, and that these could in turn prove to be of use to both the HCI and music technology communities.

Chapter 4

Understanding the Target User

As a starting point for our development of a responsive environment for distributed performance, we decided to adopt a traditional user-centered design model. Such an approach, as described earlier in Section 2.4.1, is based on an early and thorough understanding of the target user, followed by iterative cycles of formal tests and design improvements. Our choice of this particular user-driven methodology was motivated in part by the broad nature of our initial research goals. While we had established the enhancement of distributed performance as our foremost objective, and created guidelines to steer our efforts, our vision for any concrete functionality at such an early stage of design was lacking at best, as we could not anticipate which specific functionality might help improve the experience of distributed performers. Furthermore, although our target users would naturally be musicians, such a demographic was still considered rather broad. As such, we also had to determine the specific *type* of musician towards which our design efforts should be tailored. Therefore, we set out to gain a thorough understanding of various types of musicians, with a specific focus on their interpersonal interactions and motivations.

4.1 User Observations

We began by gathering extended "fly-on-the-wall" style video footage of musicians playing together in a relaxed environment. The participants we worked with and filmed varied in terms of expertise and the length of time they had been practicing music with each other, as seen in Table 4.1. Note that the term "Active" in the last column refers to the length of time for which all members had been playing together for when our observations began.

Band	Players	Genre	Song Type	Instruments	Sessions Type	Active
1	3	Rock	Covers	Electric Guitar Acoustic Guitar Electric Bass	Jam	A Few Weeks
2	4	Metal	Originals	Electric Guitar Electric Bass Drums Vocals	Rehearsal	Two Years
3	3	Jazz	Covers	Electric Guitar Electric Bass Keyboard	Jam	Never
4	3	Jazz	Covers	Electric Guitar Keyboard Vocals	Jam	A Few Weeks
5	4	Rock	Originals	Electric Guitar (lead) and Vocals Electric Guitar (rhythm) Electric Bass Drums	Rehearsal	One Year

Table 4.1: Description of Participating Bands.

4.1.1 What we noted

The following list provides a brief summary of the observations made while working with each of the participating bands:

- **Band 1:** The musicians did not move about very much, and spent a large proportion of their time staring at each other's instruments. We noted a high level of verbal communication between band members even while playing, as they seemed at times to be unsure of how to proceed. The electric guitar player, who seemed to have previous experience playing many of the selected covers, took on the role of an instructor, advising the other musicians on how to play particular progressions.

- **Band 2:** The musicians were far more comfortable with one another in comparison to Band 1, and this was particularly notable in their body language. They made full use of their surrounding space, with the bassist and guitarist often moving closer to one another. In many instances, the bassist and guitarist would assume identical poses while facing each other, playing their instrument with one foot forward and their torsos leaning backwards. During most of the instrumental parts, the vocalist would shy away to the side, and look at her notes while the other musicians played on. When she was due to begin singing once more, she would return to a close formation with the bassist and guitarist.

- **Band 3:** All three musicians chose to remain seated throughout the sessions and, to ensure that our observations remained ecologically valid, we did not pressure them to stand. Instead, we decided to focus on interaction patterns more subtle in nature than full body motion. We noted that the bassist and guitarist would often spend time looking at one another, smiling and bobbing their heads in unison. The keyboardist, on the other hand, seemed rather unsure of herself, and would often withdraw her hand from the keys when the other two musicians began their solos. She would then slowly re-introduce one hand, until she gained enough confidence to play with both. Interestingly, since the structure of the jazz songs they chose dictated that they each take turns playing 8-measure solos, the guitarist and bassist would always look up in anticipation to the keyboardist when it was her time to start.

- **Band 4**: The vocalist chose to stand, while the guitarist and bassist were seated. During the instrumental parts, the vocalist seemed rather shy and would either watch the two musicians play at length, or flip through her notebook of lyrics, much like the singer in Band 2.

- **Band 5**: The musicians seemed very comfortable in each other's presence, even discussing current events while absentmindedly playing their instruments in harmony during warm-up. Their tight rehearsal space, unfortunately, did not give them much room to move about. However, the two guitarists would

> often turn to face one another directly, leaning their bodies closer and bobbing their heads in unison. The bassist, on the other hand, seemed more engaged in the music itself, often closing his eyes or staring directly at the ground, but still moving his body along with the rhythm.

As a general pattern, musicians who had performed together for longer periods tended to interact with one another through more physically pronounced movements. Within these ensembles, for instance, one musician wishing to devote more attention to a band mate would likely move towards him, which typically also signalled a desire to "groove" together. In such cases, both musicians would also commonly assume similar body postures. On the other hand, interactions among ensembles who had played together more frequently, although mostly verbal, also included head turns and sustained glances, as the musicians had a tendency to observe one another directly while remaining fixed in place. Finally, we noted that, for all ensembles, adjusting the overall mix mid-session was often a cumbersome task. Typically, all levels were selected before the start of a performance, and any desired adjustments could only be undertaken between songs, after having been discussed amongst and agreed upon by all band members.

4.1.2 What the conversation analysts noted

Inspired by Corness and Schiphorst's view that "a performance, even when non-verbal, may be seen as a form of discourse or conversation" [55], and the suggestions made by Healey et al. of parallels between informal musical and conversational interaction [93], we presented footage from the jam session held by Band 3 to a group of conversation analysts led by Dr. Karola Pitsch at the Universitát Bielefeld. Our hope was to gain some insight from a different perspective into our target users, and see how interaction in performance compared to that during conversation, a question we had considered through our previous work on the Videoconferencing Privacy System. The conversation analysts were very intrigued by the problem presented, expressing that they rarely had the opportunity to analyze extensive non-verbal interaction, especially within the context of musical performance. They began by noting that, perhaps unsurprisingly, all interaction during the performance could only be examined within a "multimodal" context, meaning that all gestures and gaze directions made little sense without the accompanying music. Next, they began to focus on turn-taking organization. One of the basic structures of conversation analysis, turn-taking is considered the strongest evidence for a claim to universality in language. More specifically, they ascribed the interactions during performance to two states, "control" and "observation", as well as two actions, "negotiation" and "repair". Within the context of jazz performance, at any given time, there is one soloist in the "control" state, while the accompanying musicians are in the "observation" state. Near the end of an 8-measure solo, all musicians begin to prepare themselves for the turn-taking process. The cues signalling the likelihood that a solo, or turn, might end constitute what is known in conversation analysis as the transition relevance phase. During that stage, the musicians engage in "negotiation" to decide who will take on the next solo. The negotiation, followed by the switch, are often marked by "repair" actions. Typically verbal, the repair phenomenon refers to attempts to clear any confusion throughout the performance, but especially that which ensues from handing the role of soloist from one musician to the next. The overall process is illustrated in Figure 4.1.

Our discussion with the language analysts proved to be quite informative. We had chosen to show the analysts footage from Band 3 specifically because we felt interaction within improvisational music might prove to be a richer problem to explore than that during tight rehearsals. What is quite interesting is that the analyst who first identified the turn-taking pattern at the end of the solos was unfamiliar with the structure of jazz. In fact, he was only later informed by a colleague that the musicians expected each solo to last for 8 measures, near the end of which they would initiate the process of changing roles. A music teacher, who had not been involved in the filming sessions, provided further confirmation of the conversation analysts' comments, explaining that "[i]f the guitar player is doing a solo which is relatively improvised, the singer won't know how long the solo will last. The guitar player will get close and usually give a specific nod or look letting the singer know he has about 3-4 beats (depending on the meter) before it's his turn to start

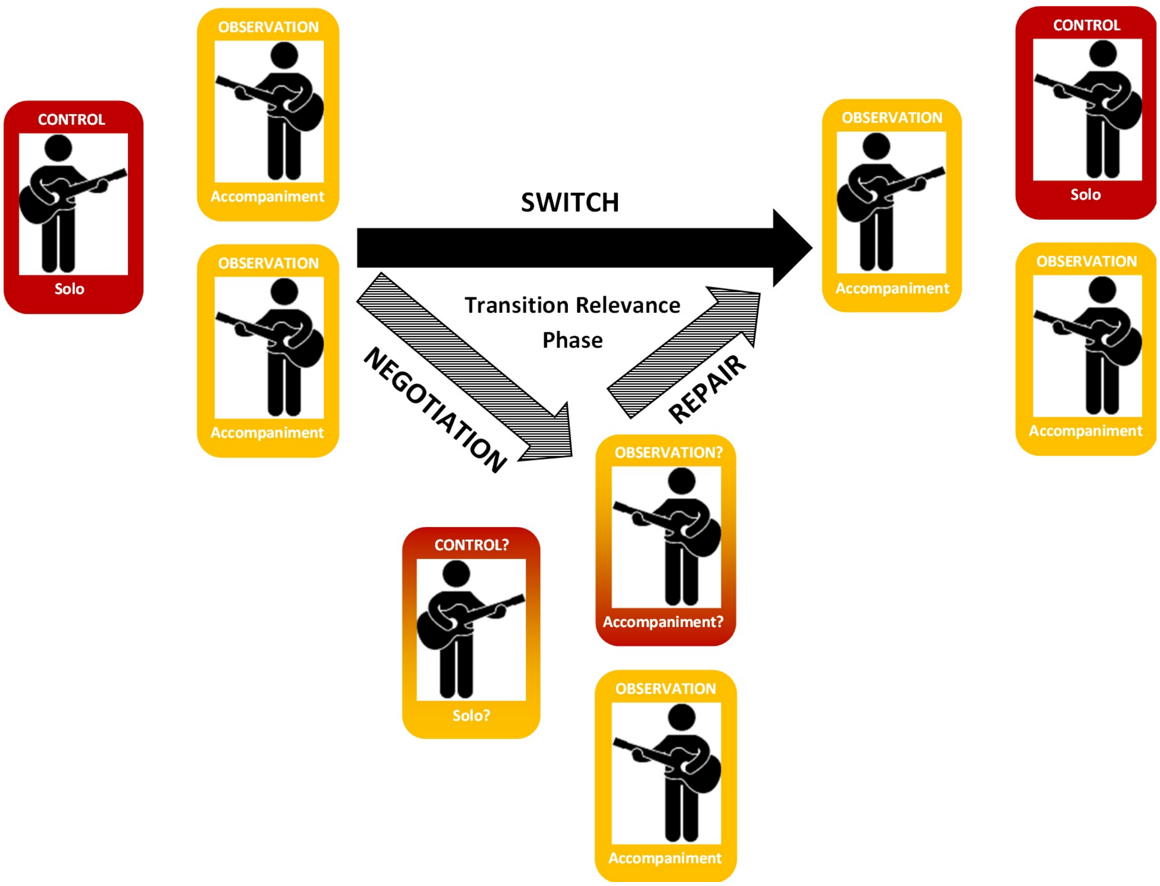

Fig. 4.1: An overview of the states and actions as described by the Conversation Analysts

up again." Furthermore, she clarified that a guitar player may go up to a singer's microphone to harmonize vocally because it enables them to tune better by hearing each other naturally, rather than waiting for digital feedback. "Band dynamics are always interesting," she added. "You will notice the singer will flash a smile to the drummer from time to time... This isn't just appreciation or love, it's letting the drummer know something in a non-verbal manner. A nod up can mean pick on [sic] the pace and a nod to the side can mean slow it down a touch. Bands will fine tune non-verbal communication but there isn't a standard."

The typology identified by the conversation analysts was extremely useful in giving an alternative context to our observations, and helped us annotate our video footage and classify the observed interactions in greater detail. Such annotation was essential to the creation of our user personas, a process that we detail in the following section.

4.2 Personas

A popular tool in user-centered design, personas are descriptions of archetypal users constructed from "well-understood, highly specific patterns of data about real people" [185]. Personas put a "face" on the design target, allowing system developers to better understand and situate their users' goals, skills and abilities. Personas can be created through the synthesis of raw data acquired from user interviews, user observation, or contextual inquiry, a middle-ground approach where users are interviewed "at work". Given that we had collected extensive performance footage, we decided to utilize our findings as a starting point for our persona definitions.

As seen in Table 4.1 above, we worked with 15 different users, who differed in genre, instrument and session type. In order to create our personas, however, we had to group musicians deemed similar enough into subsets that could be represented as accurately as possible through one, all-encompassing description. Therefore, we began by investigating which factors most accurately connect various types of musicians: was it the genre of music they preferred? Was it the instruments they played? Or was it their motivations for engaging in musical performance? In order to derive a rigorous, rather than speculative, answer to such questions, we adapted elements from Young's methodology for deriving Task-Based Audience Segments [195]. Young defines such audience segments as "groups of people who do similar things". As this description also reflects the very foundation behind personas, her methodology proved to be, with some modifications, an effective tool for creating such user profiles. Namely, our approach encompassed the following four steps:

1. **List Distinguishing Behaviours:** Create a detailed overview of all the ways many types of individuals might behave.

2. **Group the Behaviours:** Examine the behaviours and group them appropriately to create a smaller, more manageable set of actions.

3. **Group the Performers:** Group various types of users according to the behaviours they exhibit or actions they perform.

4. **Create the Personas:** Create a profile that best describes each group of performers.

Our first step was, therefore, to examine our annotated footage and create an exhaustive list of all actions undertaken by our participating musicians.

4.2.1 Listing the Distinguishing Behaviours

While Section 4.1.1 gives a brief overview of behaviours we found interesting or surprising, the accurate design of a persona must be informed by *all* aspects of user behaviour. Therefore, we created an exhaustive list of all the behaviours we noted during the user observation phase, as summarized in Table 4.2. The granularity of the actions within the list reflects a focus on performer-performer rather than performer-instrument interactions. Thus, for example, while labels such as "move closer to another musician" or "turn to face another musician" define very specific behaviours, broader descriptions such as "play a solo" encompass any actions used to produce sound using the instrument itself, such as fretting a chord or hitting a cymbal.

Adjust effect dial	Adjust volume dial
Ask another musician to play a note to better tune own instrument	Ask for clarification
Bob head alone	Bob head in unison with another musician
Change instruments	Chat with another musician
Check levels	Close eyes
Count down beat	Decide which song to play
Discuss aspects of the song that need change	Do the windmill
Double up on a microphone	Gaze at another musician
Gaze at another musician's instrument	Gaze at own instrument
Gaze at the ground	Look through a notebook or music sheets
Move closer to another musician	Play absentmindedly while others tune their instrument
Play accompaniment to a solo	Play a solo
Play other parts of a song	Shift one leg forward
Show another musician how to play a particular segment	Stand off to the side
Stop in the middle of a song	Talk about setlist
Take a break	Tune the instrument
Turn to face another musician	Use a pedal

Table 4.2: A List of all noted user behaviour.

4.2.2 Grouping the Behaviours

The next step towards creating our personas was to group the items in Table 4.3 in terms of affinity. Given our interest in exploring the social and communicative interactions employed by musicians, we devised the following taxonomy for categorizing our list of observed user behaviours:

1. **Playing Music:** Any actions necessary to the production of sound or preparation of instruments

2. **Social Behaviour:** Actions that involve two-way interaction or communication between musicians

3. **"Semi-Social" Behaviour:** Actions that involve one-way interactions between players, or that have the potential to engage or draw the attention of another musician

4. **Solitary Behaviour:** Actions that involve only one player detached from the others

Category	**Behaviours**
Playing Music	Adjust effect dial Adjust volume dial Change instruments Check levels Count down beat Look through a notebook or music sheets Play accompaniment to a solo Play a solo Play other parts of a song Show another musician how to play a particular segment Stop in the middle of a song Tune the instrument Use a pedal

Social Behaviour	Ask another musician to play a note to tune own instrument Ask for clarification Bob head in unison Chat with another musician Count down beat (1,2,3,4!) Decide which song to play Discuss aspects of a song that need change Double up on a microphone Gaze at another musician Move closer to another musician Play accompaniment to a solo Play a solo Show another musician how to play a particular segment Take a break Talk about setlist
"Semi-Social" Behaviour"	Do the windmill Gaze at another musician Gaze at another musician's instrument Move closer to another musician Shift one leg forward Stop in the middle of a song Take a break
Solitary Behaviour	Bob head alone Close eyes Gaze at own instrument Gaze at the ground Look through a notebook or music sheet Play absentmindedly while others tune their instruments Shift one leg forward Stand off to the side Take a break

Table 4.3: Musician behaviours organized by categories.

Behaviours grouped according to our categorization scheme can be seen in Table 4.3. Examining the results, however, revealed that several tasks appeared in multiple categories, without necessarily belonging to one more than the others. Furthermore, we noted that the social merit of many actions might be better judged on a continuous scale rather than a discrete one. Finally, not all tasks in the "Playing music" category were strictly necessary to the production of sound, as a number of them could be considered rather ancillary in nature. Therefore, we decided that a two-axis organization, as seen in Figure 4.2, might produce a more meaningful task configuration. The horizontal axis indicates a task's level of sociability, from "solitary" at one extreme, to "social" at the other. The vertical axis, on the other hand, indicates a task's capacity towards producing sound, from "ancillary" to "necessary". Tasks were arranged around the axes in a relative rather than exact manner, in accordance with our observations, so as to illustrate more clearly the relationship between the various behaviours in terms of both their social and musical implications. In turn, this helps provide a more accurate overview of the similarities between user tasks.

Items appearing in clusters on our two-axis diagram were subsequently grouped together, resulting in the list presented in Table 4.4. Each group was assigned a descriptive label. Such labels, in turn, formed a smaller set of user tasks that helped simplify the remainder of the persona-building process.

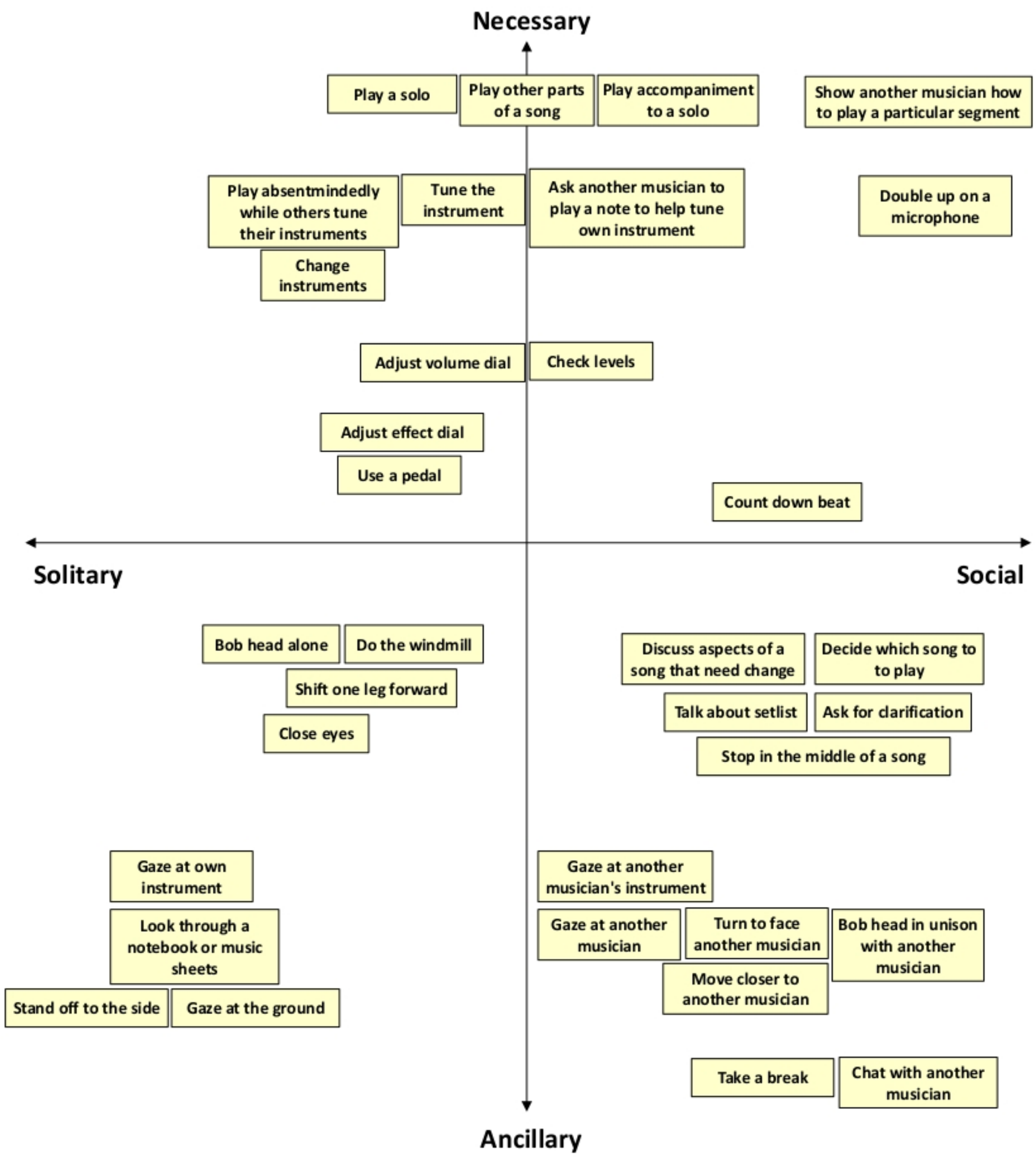

Fig. 4.2: Musician tasks represented in a two-axis organization.

4.2.3 Grouping the Performers

This step is an exercise in pattern matching. Namely, we had to match all the users we observed to the grouped actions listed in 4.4. Therefore, we began by listing the tasks down the leftmost column, and the types of performer across the top of a new table, as seen in Figure 4.3. If a user was observed performing one of the listed actions in our collected footage, we place an 'x' in the applicable cell. Note that the vocalist and lead guitarist from Band 5 is treated as two different users, "Rock Vocalist (originals)" and "Rock Lead Guitarist (originals). Given that he plays two different roles in his band, we hoped that separating them would allow for a more accurate grouping of the users. Subsequently, we examined the resulting table, ignoring the leftmost column, and focused instead on the rows of x's. Rows that were entirely checked off were then removed, since they represented universal actions, and were therefore not useful in helping us discern between different groups of users. After grouping rows with similar patterns while keeping all columns in the same order, we began looking for larger blocks of x's. The resulting configuration can be seen in Figure 4.4, where coloured blocks represent sets of users matched through behaviour similarity. As Young explains, the patterns need not be exact: there is often more than one way of grouping the x's into blocks, and outliers are common. Therefore, one must always review the resulting groups of users, and convince oneself that they really do make sense [195]. In our case, the performers comprising each block exhibited enough similarities—by role, genre, session type and/or instrument comparability—that we were satisfied with our pattern matching. Therefore, we proceeded to create personas that best depict each group.

4.2.4 Creating the Personas

After grouping our users according to similarities in their behaviours, we created persona profiles that best described the musicians who formed each group. By including information regarding the motivations, expertise and sociability of each type of user, such profiles help paint a richer picture of our target users than the simple classification by instrument and genre used at the very beginning of our observation

Group Labels	Behaviours
Adjust volumes	Adjust volume dial Check levels
Adjust effects	Adjust effect dial Use a pedal
Count down beat	Count down beat
Discuss performance	Ask for clarification Decide which song to play Discuss aspects of a song that need change Stop in the middle of a song Talk about setlist
Double up on a microphone	Double up on a microphone
Feel the music	Bob head alone Close eyes Do the windmill Shift one leg forward
Interact with others mid-performance	Bob head in unison with another musician Gaze at another musician Gaze at another musician's instrument Move closer to another musician Turn to face another musician
One-sided behaviours	Gaze at own instrument Gaze at the ground Look through a notebook or music sheets Stand off to the side
Play a song	Play accompaniment to a solo Play a solo Play other parts of a song
Prepare instruments	Ask another musician to play a note to help tune own instrument Change instruments Play absentmindedly while others tune their instruments Tune the instrument
Show another musician how to play a particular segment	Show another musician how to play a particular segment

Table 4.4: Musician behaviours grouped by similarity, according to the clusters appearing in the two-axis organization.

	Heavy Metal Guitarist	Heavy Metal Bassist	Heavy Metal Drummer	Heavy Metal Vocalist	Jazz Vocalist	Jazz Guitarist	Jazz Bassist	Jazz Keyboardist	Rock Vocalist (originals)	Rock Lead Guitarist (originals)	Rock Rhythm Guitarist (originals)	Rock Bassist (originals)	Rock Drummer (originals)	Rock Electric Guitarist (covers)	Rock Bassist (covers)	Rock Acoustic Guitarist (covers)
Adjust effects	x	x								x	x	x				
Adjust volumes	x	x		x	x	x	x	x	x	x	x	x		x	x	
Count down beat	x	x	x						x				x			
Discuss performance	x	x	x	x	x	x	x	x						x	x	x
Double up on a microphone	x	x														
Feel the music	x	x	x	x	x	x	x		x	x	x	x				
Interact with others mid-performance	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Play a song	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Prepare instruments	x	x		x	x	x	x	x	x	x	x	x		x	x	x
Show another musician how to play a particular segment	x	x				x	x	x						x	x	x
Socialize with others	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Solitary behaviours					x	x										

Fig. 4.3: Blocks of x's with similar patterns in groups.

	Heavy Metal Guitarist	Heavy Metal Bassist	Heavy Metal Drummer	Heavy Metal Vocalist	Jazz Vocalist	Jazz Guitarist	Jazz Bassist	Jazz Keyboardist	Rock Vocalist (originals)	Rock Lead Guitarist (originals)	Rock Rhythm Guitarist (originals)	Rock Bassist (originals)	Rock Drummer (originals)	Rock Electric Guitarist (covers)	Rock Bassist (covers)	Rock Acoustic Guitarist (covers)
Solitary behaviours					x	x										
Count down beat	x	x	x						x				x			
Double up on microphone	x	x														
Show another musician how to play a particular segment	x	x				x	x	x						x	x	x
Discuss performance	x	x	x	x	x	x	x	x						x	x	x
Prepare instruments	x	x		x	x	x	x	x	x	x	x	x		x	x	x
Adjust volumes	x	x		x	x	x	x	x	x	x	x	x		x	x	
Feel the music	x	x	x	x	x	x	x		x	x	x	x				
Adjust effects	x	x								x	x	x				

Fig. 4.4: Blocks of x's with similar patterns in groups.

process. The persona descriptions can be seen in Figures 4.5 through 4.9.

4.2.5 Choosing our Target User

By allowing us to downsize our initial group of 15 musicians into five *types* of users, the process of creating personas helped us refine our understanding of the musician as a target user. It also resulted in brief profiles we could keep front and center throughout the design process, especially when the users themselves could not be directly available. Nonetheless, we knew that tailoring a specialized system to the specific needs of five different user archetypes could prove to be too expansive a task. Therefore, we decided it would be best to select one type of user on which to focus. To make such a selection, it was important for us to consider not only who would benefit most from our system, but also would be most interested in experimenting with a novel environment and new contexts of performance. It quickly became apparent that expert musicians were perhaps best suited to such criteria, as they would

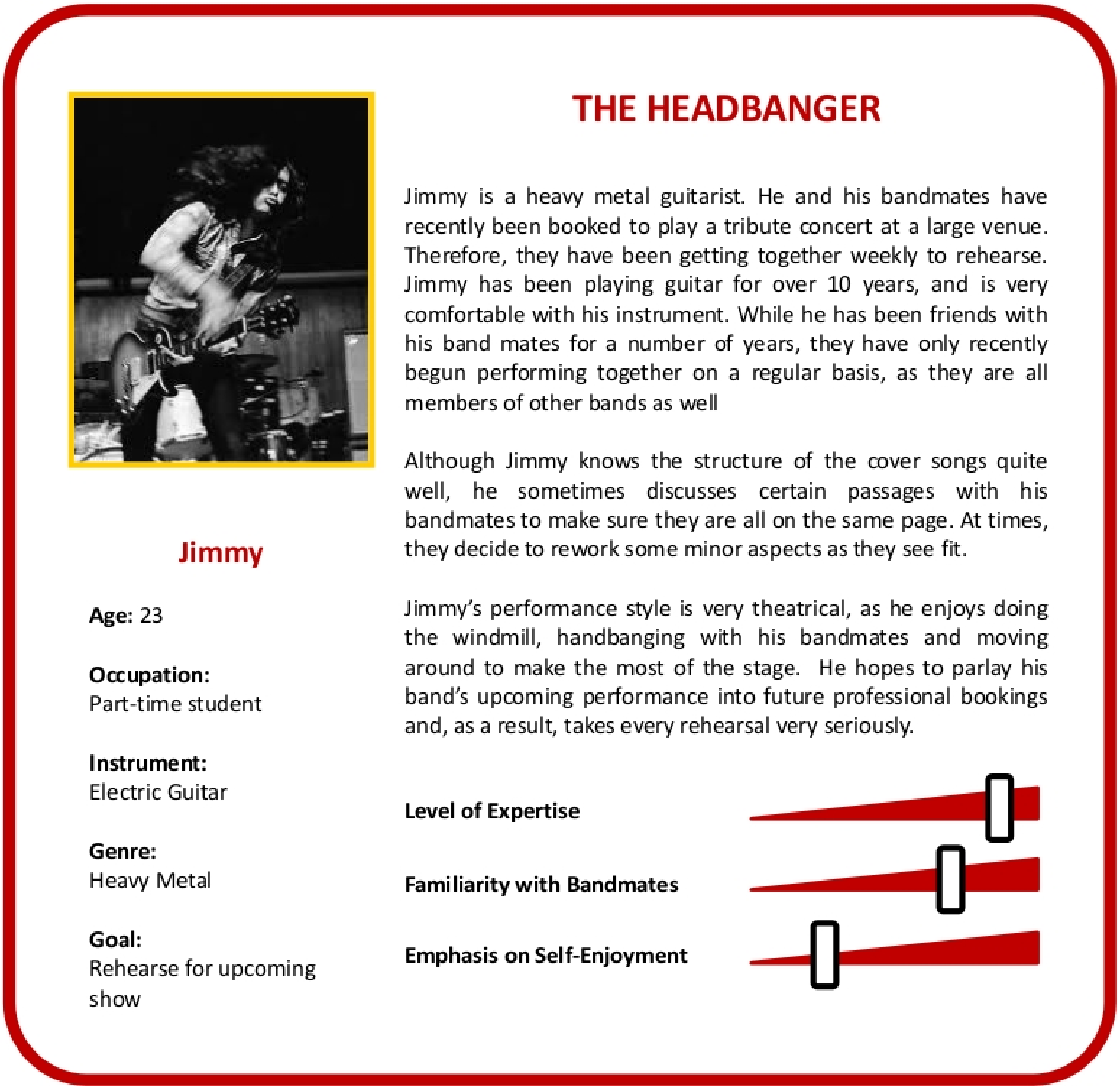

Fig. 4.5: Persona profile of "The Headbanger.

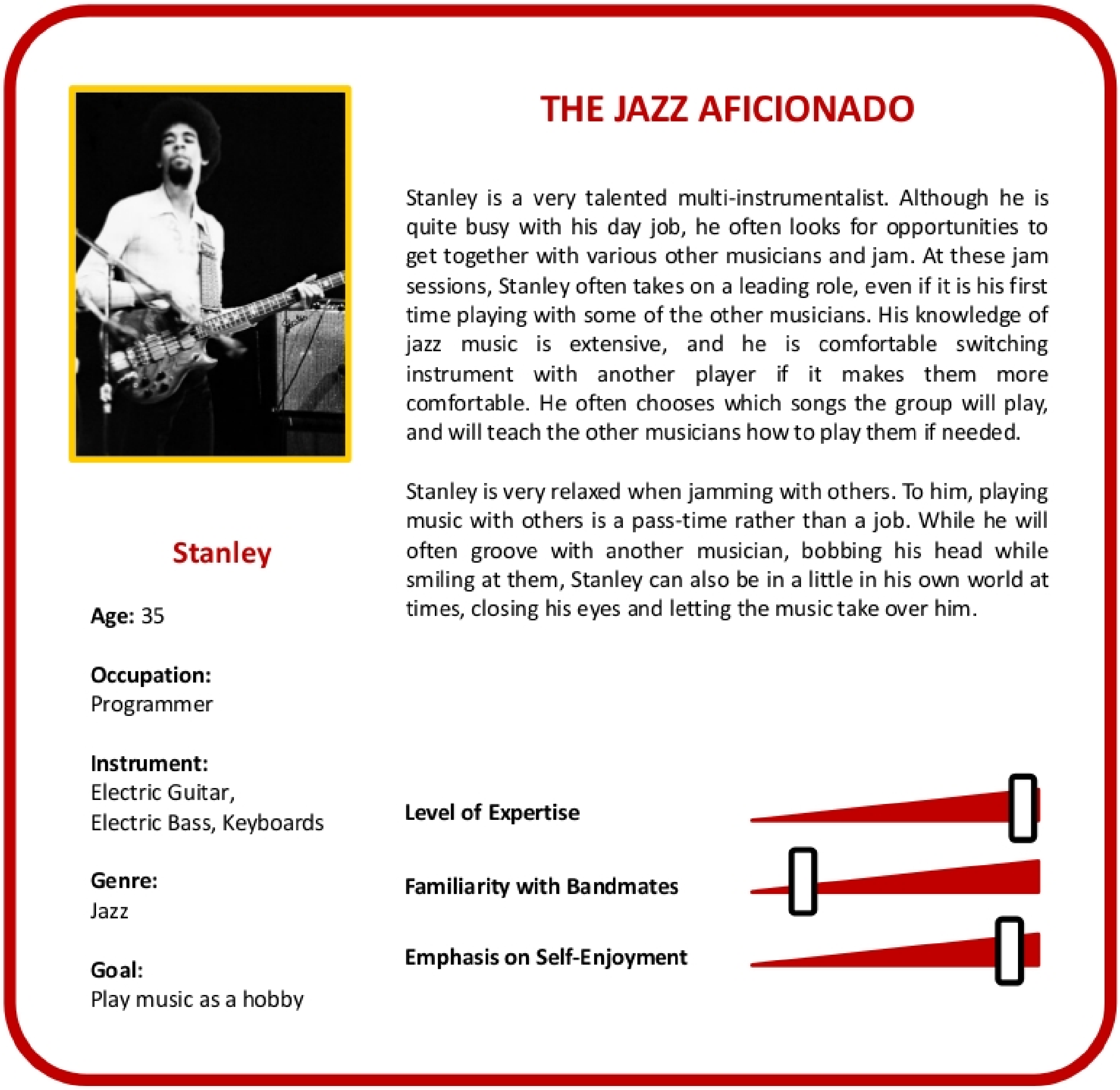

Fig. 4.6: Persona profile of "The Jazz Aficionado".

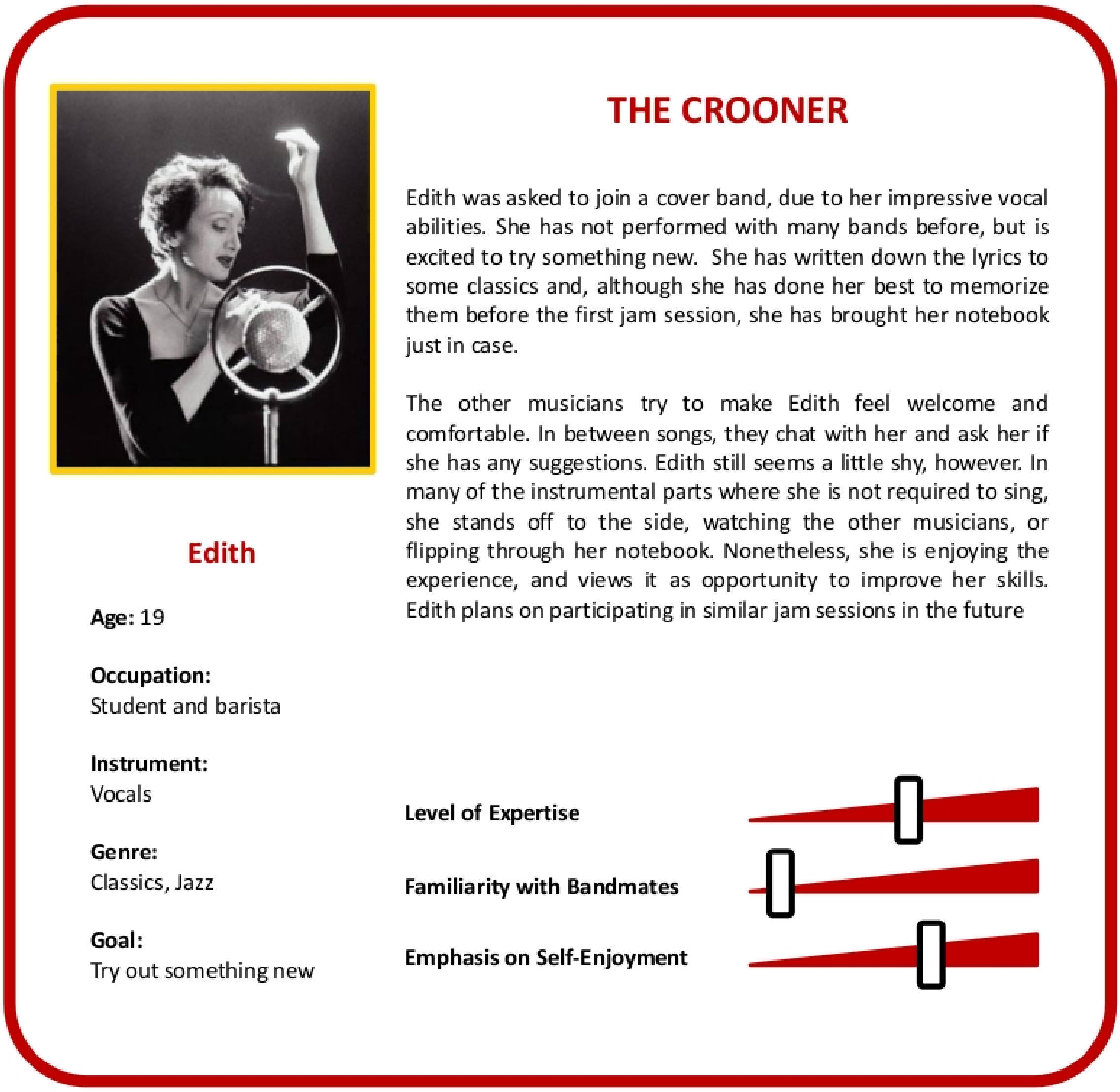

Fig. 4.7: Persona profile of “The Crooner”.

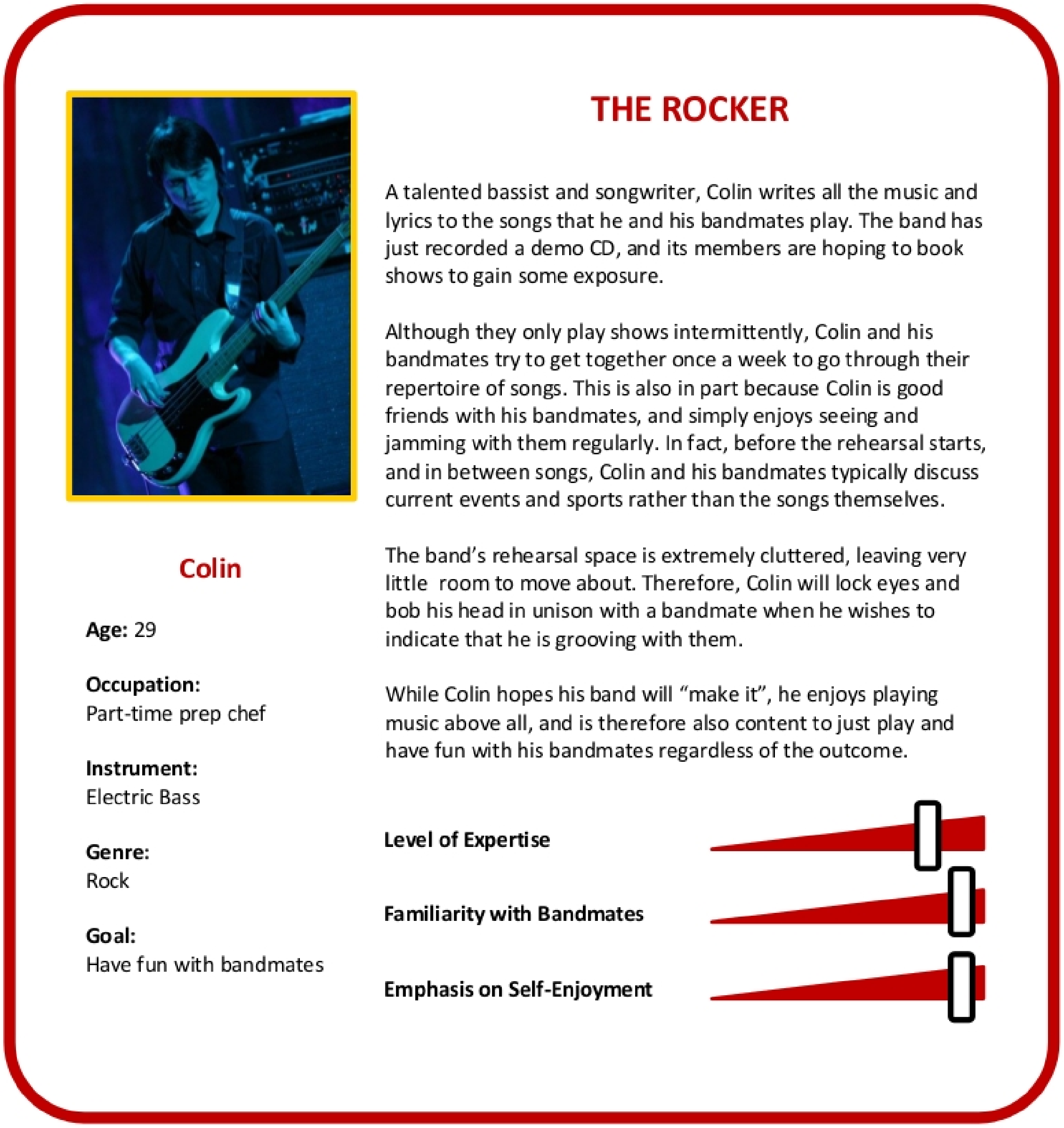

Fig. 4.8: Persona Profile of the Serious Rocker.

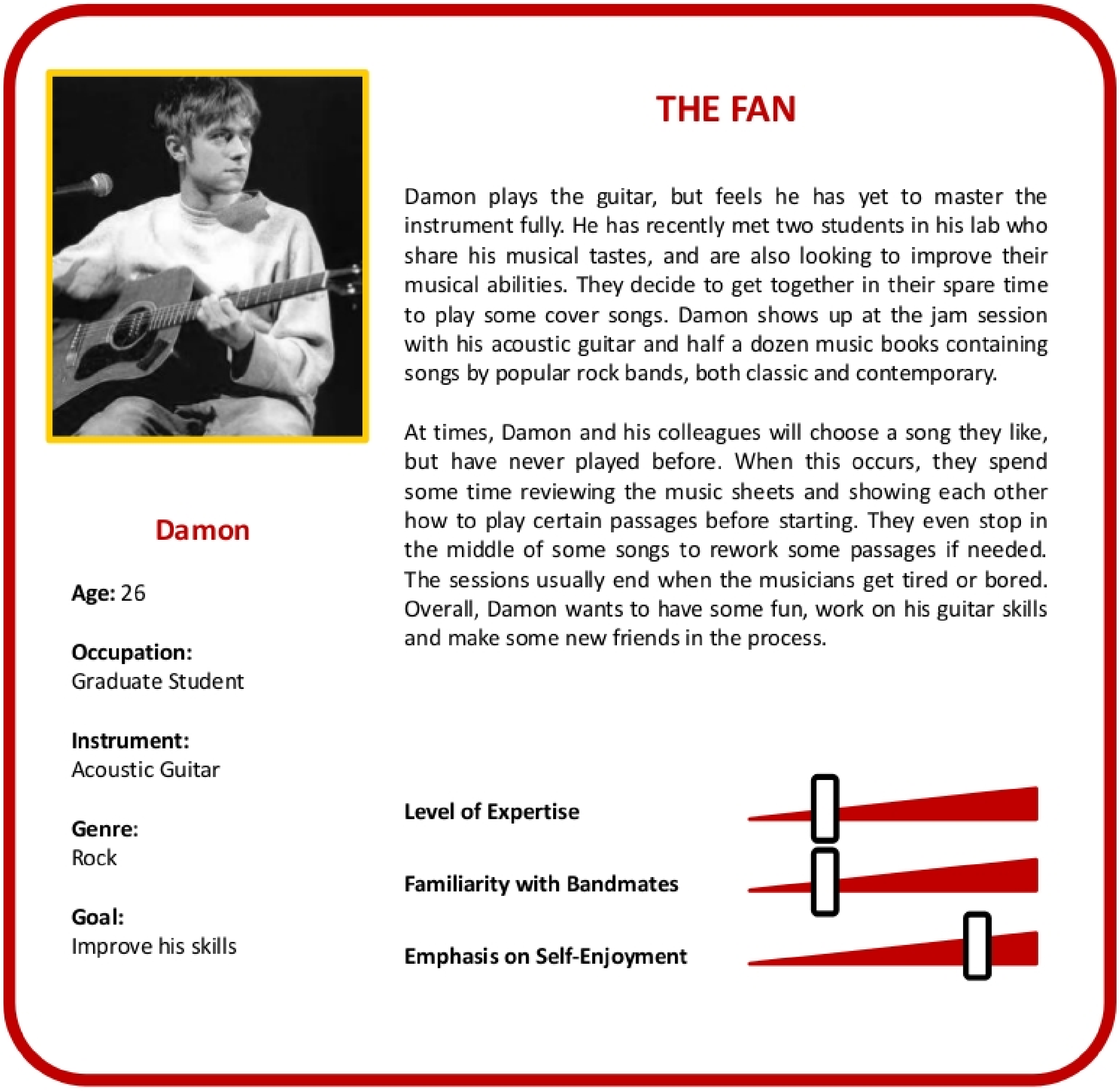

Fig. 4.9: Persona profile of "The Fan".

not be burdened with having to master their instruments while learning to interact with a new system. Furthermore, musicians showing a preference for relaxed and improvisational performances, in contrast with those who play professionally, would be more open towards exploring new interfaces. Finally, as explained in Section 4.1.1 above, we noted that the level of familiarity between musicians strongly affected their interpersonal interactions. Given our interest in examining the effects of displacement on such interactions, and measuring the extent to which augmenting distributed performance would further affect them, we deemed it important to work with users who were already familiar and comfortable with one another. By closely comparing our different personas profiles, as seen in Figure 4.10, we concluded that "The Rocker" best exemplified the type of user who stood to benefit most from our efforts. Thus, from that point on, we worked exclusively with users who fit that profile, and strove to tailor our design approach to meet their expectations.

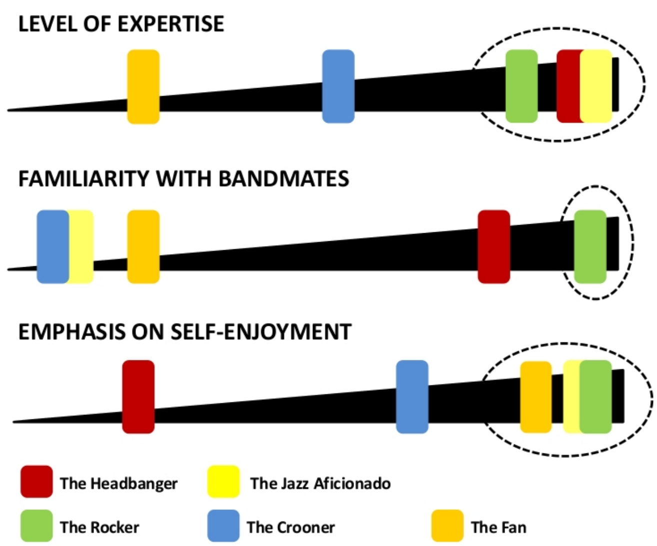

Fig. 4.10: A visual comparison of our various personas in terms of expertise, familiarity with others and emphasis on self-enjoyment. The Rocker is the only persona to score high in all three categories.

Chapter 5

Early Prototypes

The design of our early prototypes was driven by the goals and guidelines detailed previously in Section 3.3. Namely, we wanted to provide musicians with greater control over their instrumental mix, capitalizing insofar as possible on interactions already familiar to them. To this end, we mined the footage we gathered for ideas and quickly noted that, typically, musicians could not easily adjust volume levels mid-performance. In fact, in a live performance, this task is typically relegated to a sound man. We also observed that, in traditional co-present settings, band dynamics commonly included musicians moving about their space, getting closer and further away from one another during various parts of a performance. This formed the basis for our first feature, titled “dynamic volume”, which regulates volume, as heard through headphones, directly on the basis of the relative positions (proximity) between musicians, allowing them to experience each other’s volumes as getting louder when they move closer to one another. We hoped, in addition, that this might encourage distributed musicians to take full advantage of their performance space as they do in co-present performance, and thereby increase their level of interaction with one another.

5.1 First Prototype

The earliest incarnation of our system was designed to test a preliminary version of the dynamic volume feature, which requires a combination of position tracking and audio processing. Given that such a system does not include any graphical elements, we recognized that a paper prototype, though quick, would not be suited to our needs. At the same time, by turning to the computer for a low-tech prototype, such as that advocated by Muller and described earlier in Section 2.4.2, we were concerned about the possibility of being sidetracked by the technical complexities of our system. Therefore, to ensure that the majority of our time and energy would remain focused on the user, rather than development, we established the following criteria:

1. **Ecological Validity:** All necessary hardware must be readily available and easy to transport. This would enable us to conduct tests in our subjects' usual rehearsal spaces, with which they are already comfortable and familiar.

2. **Rapid Audio Prototyping:** The chosen development platform must include libraries for handling real-time audio processing, in order to ensure timely prototyping and debugging of simple functions.

3. **Low Overhead User Tracking:** By "overhead", we refer to time and effort spent, as we still wanted to enjoy the benefits that paper prototyping affords, namely, "maximum feedback for minimal effort" [172]. Therefore, we needed to identify a user-tracking platform that could provide sufficiently robust data for our needs, while still remaining portable to comply with our "Ecological Validity" criteria.

In light of these guidelines, we decided to develop our prototype using Pure Data and the reacTIVision[1] engine.

[1] `http://reactivision.sourceforge.net/`

Pure Data

Pure Data is a visual programming language developed by Puckette. An open source project, Pure Data was designed to facilitate the creation and exchange of interactive computer music. Pure Data users have access to a large number of objects that can be "patched" together to model audio flow and control. Through these objects, sounds can be synthesized from scratch, or captured and manipulated, then played back. Also notable for our purposes is that Pure Data can easily communicate with other applications and machines through the Open Sound Control (OSC) protocol. Furthermore, as described earlier in Section 3.1, the `~nstream` external for Pure Data allows for low-latency transmission of uncompressed audio, a feature that could be highly desirable when developing prototypes for the distributed context.

ReacTIVision

ReacTIVision is an open source, cross-platform computer vision framework developed by Kaltenbrunner and Bencina at the Music Technology Group of the Universitat Pompeu Fabra. It was originally designed as a component of the Reactable, a tangible modular synthesizer. More specifically, the reacTIVision engine can identify fiducial markers and detect their position. All identification and position information is subsequently sent through the Tangible User Interface Object (TUIO) protocol, an open framework that defines a common protocol for tangible multi-touch surfaces. As a result, the reacTIVision engine can provide fast and robust tracking of physical objects. Furthermore, a version of the reacTIVision TUIO client was developed specifically for Pure Data. Therefore, with minimal effort, reacTIVision could easily be integrated into our prototype design.

5.1.1 System Configuration

The configuration of our preliminary prototype can be seen in Figure 5.1. First, a Logitech HD Pro C910 wide angle USB webcam was attached to the ceiling. Each musician was then asked to wear a pair of Sennheiser HD Pro 280 closed headphones, to which we attached a fiducial marker that could be seen by the camera, and thereby

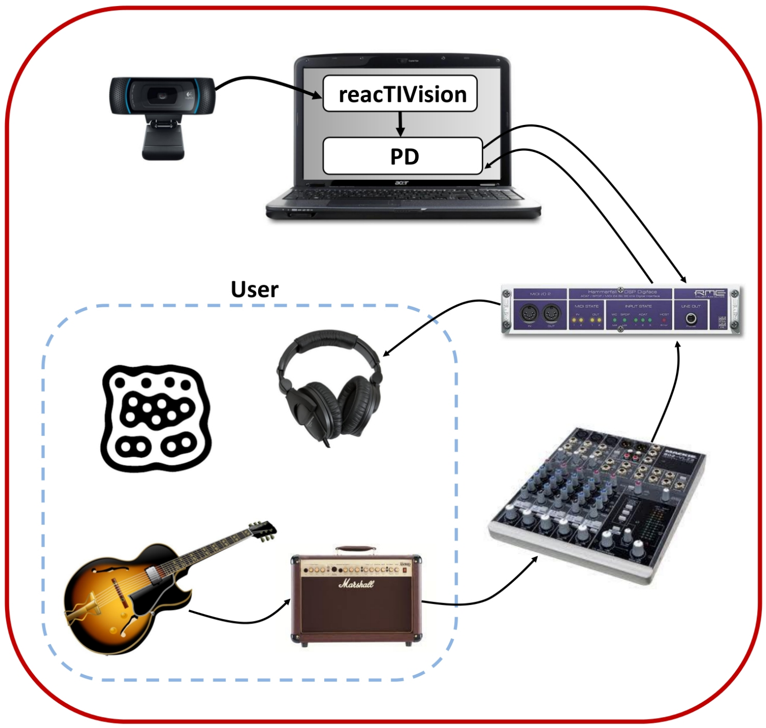

Fig. 5.1: System configuration for our first prototype. The dashed box titled 'user' represents components given to each participating musician.

tracked by reacTIVision. In turn, all identification and position information was sent to Pure Data through the TUIO client, then used to calculate the distance between musicians. Each instrument was plugged into an amplifier, and all amplifiers were in turn connected to a Mackie 1202-VLZ3 4-channel mixer. The output of each channel was subsequently fed into an RME Hammerfall Multiface II, a device that can transfer analog and digital audio directly to a computer through the accompanying RME HDSP Cardbus. All audio streams were then processed, and individualized mixes were created in Pure Data before being driven to each musician's headphones back through the Hammerfall Multiface's output channels. If two musicians moved closer to one another until the distance between them fell below a pre-determined threshold (chosen in accordance with the size of the rehearsal space), they would perceive each other's volumes as gradually increasing in volume. Thereby, performers could dynamically create and adjust their individual soundscapes.

5.1.2 User Feedback

We conducted some informal tests using this initial setup, during which musicians were asked to explore the system by moving about while playing their instruments. Unfortunately, a number of problems quickly became apparent. First, by placing the fiducial markers on our users' heads, and therefore closer to the ceiling camera, we significantly reduced the effective area where they could be seen. Thus, reacTIVision often lost track of the musicians, even if their feet or torsos could still be seen by the camera. This meant that all musicians had to stand very closely to one another in order to be tracked, making it very difficult for them to increase their proximity and experience the effects of dynamic volume any further. On the other hand, when placed on the musicians' shoes, the markers were too small to be seen by the camera, and were often obscured by their bodies. Furthermore, some musicians found the level of latency they experienced, which occurred as a result of the rather large number of hardware components that handled the audio stream, to be problematic. Finally, the levels of noise that seeped into the final audio mixes were deemed less than ideal. Additional tests indicated that such noise resulted from the large number of cables used, along with the real-time conversion of audio streams from analog to digital by Pure Data. Although the musicians reported being able to perform despite the noise, it proved to be a source of irritation. In light of these issues, we were unable to evaluate this particular prototype from a usability standpoint as we had hoped. Therefore, we were forced to re-examine our software and hardware configurations before proceeding with any further developments.

5.2 Second Prototype

For our second prototype, we decided to forgo Pure Data in favour of Supercollider,[2] and to develop a simple tracking algorithm using existing functions from the Open Source Computer Vision (OpenCV)[3] library. Such tools, as described below, still allowed us to adhere to the lo-tech prototyping criteria established earlier.

SuperCollider

SuperCollider is an open source environment and programming language for real-time audio synthesis and algorithmic composition. Each instance of SuperCollider initiates a server that can communicate to local and remote clients via OSC messages, making SuperCollider ideal for distributed audio. In addition, SuperCollider is a dynamic programming language, allowing for quick modular testing and debugging. Most importantly, preliminary tests showed that SuperCollider exhibited levels of noise and latency far lower than those experienced with Pure Data, making it a more suitable framework for our audio processing requirements.

Colour Tracking

Placing a camera on the ceiling meant that fiducial markers could only be tracked if placed on horizontal planes or, in other words, on the musicians' heads, shoulders or feet. However, as described earlier, positioning markers too close to the camera

[2] `http://supercollider.sourceforge.net/`
[3] `http://opencv.willowgarage.com/wiki/`

significantly reduced the area in which they could be seen, while placing them too close to the ground rendered them too small for detection, and often completely obscured by the musicians' bodies. We therefore decided to instead detect different colours worn by our participants. An example of what Bongers refers to as "passive beacons" [23], a standard t-shirt covering a user's shoulders, upper arms and torso creates a large target with several planes that can effectively be seen by the camera. In turn, this could vastly increase the overall area in which the musicians could interact with one another while still being tracked by our system.

Our colour detection algorithm assumes that only one object of each colour targeted for detection is seen by the camera at any given time. Written in C++, it utilizes a number of functions from the OpenCV library in the following manner:

- A red, green and blue (RGB) frame, I_{RGB}, is captured by the ceiling camera. Within this image, we are attempting to detect an object of uniform colour with a hue, saturation and value defined to be within the ranges $h_{min} \leq h \leq h_{max}$, $s_{min} \leq s \leq s_{max}$ and $v_{min} \leq v \leq v_{max}$, respectively.

- The image must first be transformed into the HSV format using the OpenCV function `cvCvtColor`, which converts an image from one colour space to another, and results in the new image I_{HSV}.

- Subsequently, we use the `cvInRange` function to create a thresholded image according to the target colour's hue, saturation and value ranges. The result is a binary image, I_{thresh}, where the detected object is white, while all other areas are black.

- The zeroeth order moment of I_{thresh}, m_0, is then calculated using the `cvMoments` function.

- The first order moment of I_{thresh} along the x-axis, m_{1x}, is calculated using the `cvGetCentralMoment` function.

- Similarly, the first order moment of I_{thresh} along the y-axis, m_{1y}, is also calculated using the `cvGetSpatialMoment` function.

- Finally, the position of the detected object is calculated as the centroid of the selected pixels, $(p_x, p_y) = (m_{1x}/m_0, m_{1y}/m_0)$.

The position values are then broadcast via OSC, and can be used by any application of our choice, such as SuperCollider.

5.2.1 System Configuration

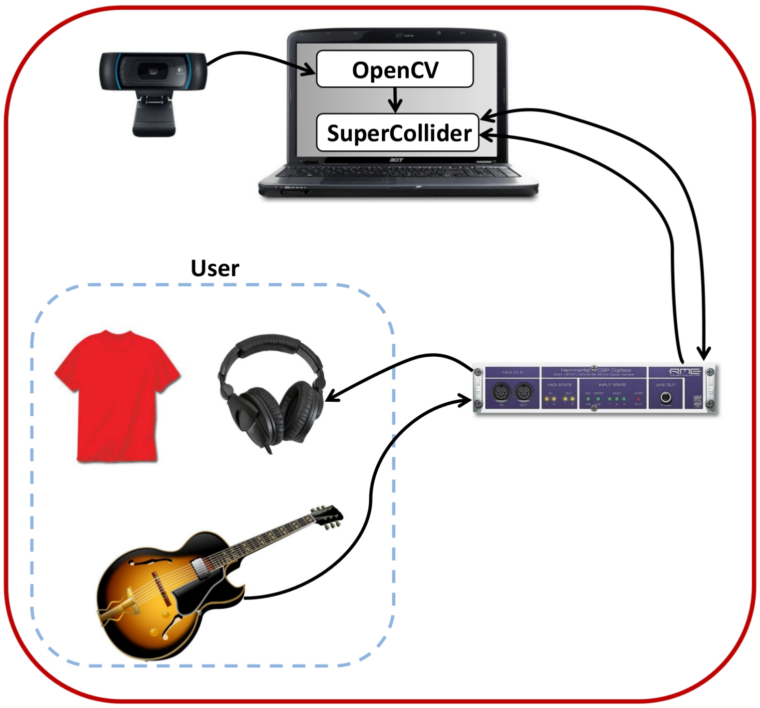

Fig. 5.2: System configuration for our second prototype. The dashed box titled 'user' represents components given to each participating musician.

In terms of hardware, our second prototype's configuration, seen in Figure 5.2, was much simpler than the first. Each participant was asked to wear a bright red, blue or green t-shirt, rather than a fiducial marker, in order to be identified and tracked by our algorithm. Furthermore, since the overall sound quality and volume achieved by connecting instruments straight to the Hammerfall Multiface proved to be surprisingly satisfactory to our musicians, the mixer and the amplifiers were not used in this configuration. This enabled us to further reduce the noise and latency levels experienced in the overall mix.

5.2.2 Preliminary User Feedback

We asked a trio consisting of a singer, keyboardist and bassist to participate in an informal test using our system, in order to gather some preliminary feedback on our new configuration and the dynamic volume feature. Throughout the session, a threshold of 250 cm was set, meaning that two musicians would begin experiencing volume changes when the distance between them dropped below that value. This threshold was chosen as an initial figure for testing purposes, and proved to be suitable. The participants were not given much detail regarding our system, only that moving around would allow them to interact with it, should they wish to do so. We watched them perform for some time until they were finally comfortable with one another. Eventually, the singer began to move closer to the bassist and stood next to him for some time. Then, she moved across the room to stand closer to the keyboardist. She continued to move back and forth between both players until, watching her, the bassist was inspired to move around as well. At the end of the performance, the singer eagerly told us that she found the system to be very exciting, and that she had in fact moved closer to each musician during their solos in order to better "focus" on what they were playing. Interestingly, the bassist noted that during traditional rehearsals, he was often frustrated at his lack of control over other musicians' volumes: the sound levels, while optimal for other players, could at times be less than ideal for him. Therefore, he indicated that dynamic volume could be beneficial in such scenarios, allowing him to create his own personalized sound mix. Finally, the musicians also reported enjoying themselves while interacting with our system.

5.3 User Interviews

Having received some positive preliminary feedback on the uses of dynamic volume, we were encouraged to continue refining and expanding our prototype. As per the mandates of user-centered design, however, the following step in the evolution of the system was to begin testing it in a more rigorous and iterative fashion. Nonetheless, a question remained: what type of performance criteria would we use when evaluating our responsive environment? Ideally, any benchmarks against which our system would be tested needed to reflect values that our target users found important. However, while our user observations and personas provided insight regarding the *what* and *how* of musical performance, we had yet to fully appreciate the *why* behind many of the musician's actions and decisions. To uncover this type of information, we decided to hold interviews with a sample of our target population. A non-leading interview format was chosen in order to minimize the possible bias that may arise from a question-and-answer style of discussion. Our first task towards conducting such interviews was to prepare prompts that would serve to gently remind us how the conversation should be steered. Overall, the areas we wanted to touch upon included the subjects' musical backgrounds, the types of rehearsals or performances in which they participated, the nature of their relationships and interactions with their bandmates and other musicians, their aspirations, and whether they had previously partaken in distributed performances or tried new musical interfaces. The next step was to recruit subjects who met our chosen persona profile, "The Rocker", at least in terms of expertise, emphasis on self-enjoyment and familiarity with bandmates. Prospective participants were asked to answer an online "screener". This consisted of a brief questionnaire that elicited information about their musical backgrounds and habits, and included a "softball" question to help the interviewer determine whether a potential subject would be a good communicator. In our case, we asked interested candidates to list any musician, alive or dead, with whom they wished they could perform, and to describe an ideal encounter with that musician. If a candidate answered the question satisfactorily and met our persona profile, they were invited back to take part in an hour-long interview. We spoke with six musicians, one female and five male, ranging in age from 18 to 42.

5.3.1 Content Analysis

To avoid bias from pre-conceived notions regarding users' needs and desires, we applied the methodologies of grounded theory and content analysis, described in Section 2.5.1, as we transcribed and analyzed the interviews. During the coding process, any quotes alluding to motivations, behaviours, preferences or values held by the musicians were assigned a descriptive tag. After all interviews had been coded, a list of all twenty-one generated tags was compiled. Subsequently, any tags considered sufficiently related were grouped, and such groups were each assigned a new encompassing label, which we will refer to as "incentives". This helped reduce the overall set of tags to a more manageable size of seven, as seen in Table 5.1. The interview transcripts were in turn labeled using the new set of tags and, as a final step, the incentives were ranked according to the overall frequency of quotes associated with each, as shown under the "Interview Analysis" column of Table 5.2 below.

5.3.2 Validation

Our next step was to check the validity of our content analysis and, in particular, the accuracy of our incentive rankings. Since we were predominantly interested in understanding the values deemed most important by musicians, we conducted a survey to determine how well our ordering would match one generated directly by musicians themselves. Our brief online survey presented participants with the list of seven incentives identified during our content analysis, and asked that they rank these values in order of importance. The survey was completed by 21 students, six female and 15 male, between the ages of 21 and 40, who roughly met our "Rocker" persona profile in terms of expertise and emphasis on self-enjoyment. None of these participants overlapped with the set of interview subjects.

Although not an exact match, the results, as shown under the "Musician Survey" column of Table 5.2, correspond reasonably well to the rankings generated through the content analysis. Most importantly, the top four values chosen by the survey respondents matched the ones produced by our content analysis, albeit in different

Group Labels	**Tags**
Creative engagement	Artistic intent Creativity Technology
Enjoyment	Enjoyment
Improving technical abilities	Challenge Improvement Motivation Technical ability
Interaction with other musicians	Body language Chemistry Collaboration Comfort Communication Interaction Movement Personalities
Professional pursuits	Professional pursuits
Live performance	Audience appreciation Energy
Self-Expression	Emotion Expressivity

Table 5.1: A list of 21 tags uncovered during content analysis, grouped in terms of similarity.

Rank	**Interview Analysis**	**Musician Survey**
1	Interaction with other musicians	Enjoyment
2	Enjoyment	Self-expression
3	Self-expression	Creative engagement
4	Creative engagement	Interaction with other musicians
5	Improving technical ability	Improving technical ability
6	Live performance	Live performance
7	Professional pursuits	Professional pursuits

Table 5.2: Incentives ranked according to the outcomes of our interview content analysis and musician survey.

order. Hix and Harston advise that the number of usability goals tested in formal experiments be kept low, citing 2-3 as an ideal figure that helps prevent testing and analysis from overwhelming developers [95]. Therefore, it was necessary to select only a few incentives from Table 5.2 to use as benchmarks against which to evaluate our system. Furthermore, only the top four incentives could, to a certain extent, be measured during an experiment, as the remaining ones required long-term monitoring of participants. Thus, we decided that any formal user experiments should help determine whether our system features would engage users creatively, allow them to better express themselves, increase their level of interaction with other musicians, and prove enjoyable to use in the process. Subsequently, our evaluation criteria were formalized according to Table 5.3. All questionnaires consisted of prompts accompanied by Likert scale ratings that ranged from 1 for “Strongly Disagree”, to 5 for “Strongly Agree”.

We note that such benchmarks are typically associated with experience rather than task-based HCI. As such, throughout the remainder of our efforts, all experiments were designed with a focus on the musicians’ overall experience with a given prototype, and did not require them to carry out specific tasks during performance.

5.4 Formal Experiment

After gaining additional insight into our target users and defining our benchmarks, we were ready to formally test our prototype. Our experiment was designed to test our responsive environment against four different benchmarks established as a result of our user interviews. Musicians were asked to choose a number of songs familiar to them, and jam for approximately half an hour, once in a traditional, non-augmented fashion, and once with our system’s dynamic volume functionality. This A/B-style test applied the adjection/intensification and separation/segmentation strategies described by Ravasio et al. for conducting qualitative research in HCI [155]. Our goal was to isolate and determine the effects of dynamic volume on the musician’s perceptions of creativity, enjoyment, self-expression and interaction. In order to maintain our experience-based approach, musicians were not required to carry out any specific

Performance Criteria	Rationale	Evaluation Method	Sample Questionnaire Prompts
Enjoyment	Musicians should enjoy themselves while interacting with our system.	In order to quantify enjoyment, we turn to ludology, where "flow" and "immersion" are often evaluated during game studies and used as indicators of a player's overall sense of pleasure. In particular, we adapted elements from IJsselsteijn's game experience questionnaire (described earlier in Section 2.5.3) to the context of musical performance, a task made feasible through the generalized nature of the questions used in the GEQ.	"I felt happy", "I felt that I was learning", "I forgot everything around me", "I lost track of time", "I felt challenged"
Creative Engagement	Our system should allow musicians to explore new grounds and help enhance their sense of creative engagement.	A questionnaire based on the most basic tenets of self-perceived creative engagement was designed to help evaluate this criteria.	"I felt that I discovered new things", "I performed differently than I usually do", "I felt inspired", "I took a risk", "I did something the others did not expect"
Self-Expression	Our system should help musicians better express their musical moods and ideas.	As with our evaluation of creative engagement, a questionnaire was created to elicit musicians' perceptions of the most basic qualities of self-expression.	"I expressed my mood musically to the others", "I expressed my feelings verbally to the others", "I expressed my ideas musically to the others", "I communicated clearly with the others", "I felt my individuality was well-preserved within the group"
Interaction Amongst Musicians	As one of our primary goals, we believe our system should help increase the level of interaction between musicians.	The game experience questionnaire includes a component on social presence, which focuses primarily on the behavioural involvement amongst musicians, and which we adapted to the context of musical performance. In addition, position data was collected throughout experiment sessions to determine how participants moved in relation to one another.	"I empathized with the others", "I felt connected to the others", "I paid close attention to the others", "I found it enjoyable to be with the others", "When the others were happy, I was happy"

Table 5.3: Description of performance criteria and their corresponding evaluation techniques.

tasks under each condition. However, as per the suggested approaches for experience evaluation described earlier in Section 2.5, the sessions were not designed to be exclusively exploratory in nature, as the musicians were instructed with an active goal of playing, as best as possible, several full songs while using our prototype. Furthermore, the musicians were instructed to voice to one another or to the test instructor any feelings or concerns they may have throughout the session. For better comparability, the musicians heard each other through closed headphones in both cases, although, naturally, the volume mix was static in the non-augmented case. Since our experience indicates that it often takes musicians a bit of time—typically a song or two—to gain some momentum and feel comfortable, or "warm up", we did not want to interrupt them between songs to switch experimental parameters. Furthermore, by avoiding changes in conditions between individual songs, we hoped to lend the experiment a more natural feel, thereby maintaining the ecological validity of the overall experience we aimed to evaluate.

5.4.1 Band 1: Results and Analysis

We first tested our system with a 4-piece band consisting of a 28-year-old female vocalist, a 22-year-old male guitarist, a 24-year-old male drum machine player, and a 22-year-old female keyboard synthesizer player. All four participants had performed together regularly in the past, and fit our "Rocker" persona profile. While our prototype had been designed to track three musicians at a time, this did not prove problematic as the synth player had no choice but to remain relatively fixed in place, due to the bulky nature of her instrument. Instead, she was assigned a static location such that we could still calculate her proximity to the other musicians as they moved about, and adjust her audio mix accordingly. In other words, she was only able to experience the dynamic volume feature passively. The drummer, on the other hand, had more liberty, as he could leave a sequence running and temporarily detach himself from his instrument to control and explore the sonic landscape surrounding him. Data was collected after each condition through the questionnaires described above, as well as in-session through audio recordings and video footage. The participants were highly encouraged to think out loud, and express any feelings or concerns they had regarding their performance.

Content Analysis

Perhaps what was most telling about our prototype were the opinions expressed by the musicians throughout the session, and later analyzed from our video footage. During the standard, non-augmented jam, the drummer and guitarist felt quite "in sync" and seemed to lead the rest of the group. They reported feeling on the same page as everyone else, and appeared relaxed, explaining that they "just let the music flow". On the other hand, the vocalist and synth player, who had more limited musical experience in comparison with their bandmates, felt out of sync, and had a hard time finding their place within the group, hesitating to contribute at times. Both reported feeling a little "out of sorts". During the augmented session with dynamic volume, the vocalist took the most advantage of the setup. We suspect that the change in condition encouraged her to begin experimenting, as she moved around the room and expressed feeling more sure of herself. The synth player also stated that she felt more at ease, although she noted that could be due in part simply to warming up. The guitarist attempted to move around, but seemed unsure of himself. The drummer expressed feeling quite frustrated at others changing the volume around him. He felt as though he could not control or "get away from" what was happening. As a result, he had difficulty finding his footing, and was unable to lead the others rhythmically as well as he would have liked. Overall, the jam session with dynamic volume sounded less cohesive in comparison to the non-augmented one.

Questionnaire Analysis

Analysis of the questionnaires completed by the participants agreed reasonably well with their expressed opinions. Figures 5.3 compares the musicians' perceptions of flow, social presence, creativity and self-expression with and without dynamic volume. The musicians' responses to questions pertinent to each of the various factors were combined to calculate values in the range of 0-1. The guitarist fared worse than the others, experiencing a sharp drop across all factors when dynamic volume was used. Similarly, with dynamic volume, the drummer experienced an increase only in his sense of social presence. The vocalist experienced little change in flow and social presence across both conditions, but felt her creativity and self-expression drop with the use of dynamic volume. The synth player, on the other hand, performed better with dynamic volume, with the exception of her sense of social presence. We suspect the vocalist and synth player were better able to cope with the unexpected changes in volume due to the more improvisatory nature of their roles within the ensemble, as opposed to the drummer and guitarist who found it more important to keep a beat.

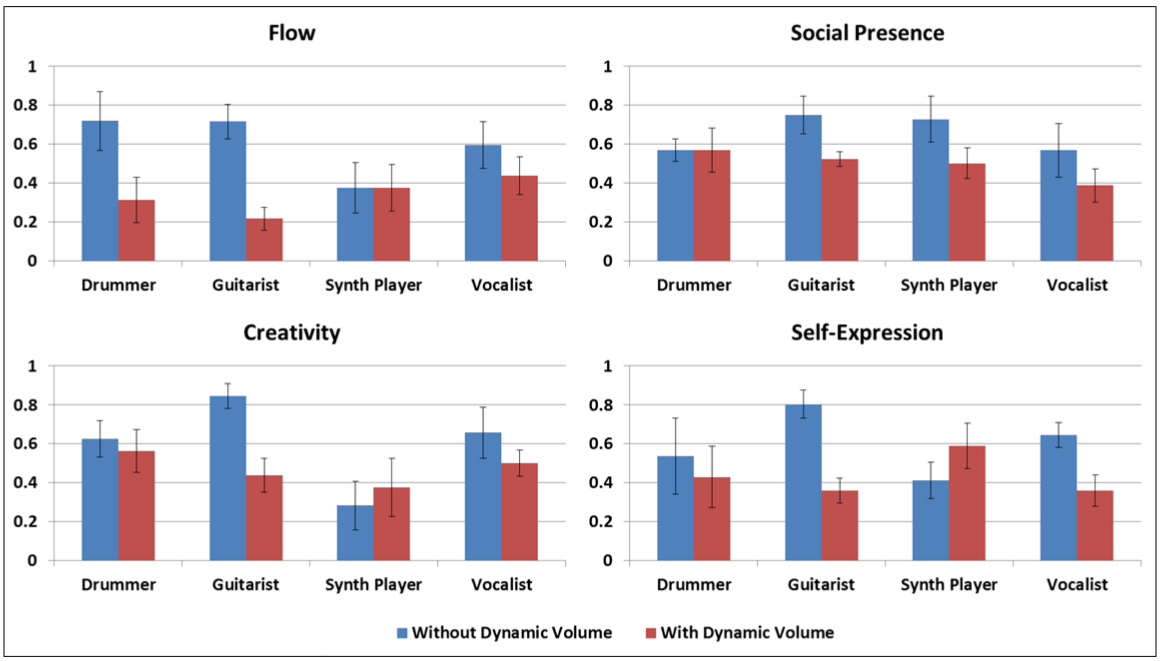

Fig. 5.3: Post-condition questionnaire results for Band 1 members. Averages are shown with standard deviations.

Data Analysis

Figure 5.4 shows the musicians' positions throughout the sessions with and without dynamic volume. As explained earlier, the synth player was assigned a static location. A comparison of both graphs shows some increase in the participants' movements when dynamic volume was used. Perhaps more interestingly, however, additional analysis of the position data helped shed some light on the negative impressions registered through the footage and questionnaires. First, we note that the field of view of the ceiling-mounted camera covered an effective area of approximately 300×400 cm, inside which the musicians had to remain in order to be tracked. Throughout the session with dynamic volume, a threshold of 250 cm was set, meaning that two musicians would begin experiencing volume changes when the distance between them dropped below that value. This threshold had previously proven successful within comparable spatial constraints imposed by the camera during our past work with another group of three musicians, as described earlier in Section 5.2.2.

When we examined the musicians' proximity to one another, however, we quickly noted that this threshold was unsuitable for a four-piece ensemble. In fact, the starting positions chosen by the musicians meant that the distance between them was already below the threshold when they began playing, and this instantly threw the drummer and guitarist off track. This also meant that they struggled to find locations where the overall mix could return to the default values they were pleased with during the session without dynamic volume. Figures 5.5 and 5.6 show overviews of the distance between the guitarist and the drummer, respectively, and the other players. As illustrated, all distances remained below the threshold, with the exception of that between the drummer and the guitarist. Only the singer moved around extensively in an attempt to experiment with the feature, much to the detriment of the guitarist and drummer, who found the resulting changes in volume to be frustrating.

Lessons Learned

As our analysis indicated, the performance with dynamic volume did not meet our established performance criteria. However, the lessons we learned through our experiment were invaluable towards the refinement of our system. For instance, the fact that the threshold used to activate the dynamic volume was appropriate for one group of musicians but not another suggested that this value should be tailored to the size of the ensemble using the system. In addition, we learned the importance of giving users "safety" positions to which they could return should they become overwhelmed by the volume changes induced by the movement of other musicians. This could be accomplished by positioning the musicians on clearly marked starting locations that allow the distances amongst them to be greater than the dynamic volume threshold. If the volume changes prove to be too overwhelming or confusing at some point, the musicians could return to their marked locations and, hence, their default audio mix. After these changes had been incorporated, we set out to evaluate our system once again.

5.4.2 Band 2: Results and Analysis

We tested our refined system with a 3-piece rock band consisting of a bassist, lead guitarist, and rhythm guitarist. The musicians were between 27 and 31 years of age, all male, and had performed together in the past, rehearsing and playing live shows regularly for nearly two years. We monitored the distances between members of the ensemble during the session without dynamic volume and found that they appeared to be comfortable at separations of approximately 230–250 cm, typically maintaining those distances between one another. Therefore, we determined 225 cm to be a reasonable threshold to set for the dynamic volume session, as it would imply an explicit move to trigger that function. Data was collected post-session through the same questionnaires described in Table 5.3, as well as in-session through video footage, audio recordings and position tracking. The participants were highly encouraged to think out loud, and express any feelings or concerns they had regarding their performance.

Content Analysis

Looking at our video footage, we were able to gain more insight into the musicians' impressions of our system. First, even though they were given a description of the dynamic volume feature before the start of the session, they were quite pleasantly surprised when they began interacting with the system. They began by moving all around the space to "get a feel" for the volume shifts. When they were more comfortable, they started taking further advantage of dynamic volume, with the rhythm guitarist and bassist, for instance, huddling around the lead guitarist as he played a solo. Finally, the rhythm guitarist commented explicitly that he had never experienced anything similar, and was quite happy to be given the opportunity to participate in our test session.

Questionnaire Analysis

Figure 5.7 compares the musicians' perceptions of flow, social presence, creativity and self-expression with and without dynamic volume. Again, the scores assigned to each factor were tabulated as averages between 0 and 1 of the musicians' responses to questions pertinent to the various facets of that factor. Overall, the dynamic volume feature fared quite well across the board, with at least two of the performers reporting an equal or improved experience with dynamic volume on all factors.

Data Analysis

As seen in Figure 5.8, during the session without dynamic volume, the musicians did not venture far from their starting positions. The only notable exception was an instant when the rhythm guitarist briefly wandered across the performance space, before returning to his original post. In contrast, however, when dynamic volume was used, all three musicians were far more adventurous, making full use of the performance space. Position data timestamps were synced with those of the video footage and audio recordings. In turn, this helped further illustrate, as seen Figure 5.9, our video footage observations of instances when the other musicians moved closer to the lead guitarist during his solos.

Lessons Learned

We found the feedback from Band 2 to be very promising. The musicians reported being quite happy to be asked to take part in the experiment, as it gave them a chance to try out a new feature that they found rather exciting. In addition, the marked increase of interplay between them strongly indicated to us that dynamic volume could potentially prove helpful in addressing the lack of sociability often seen in distributed performance.

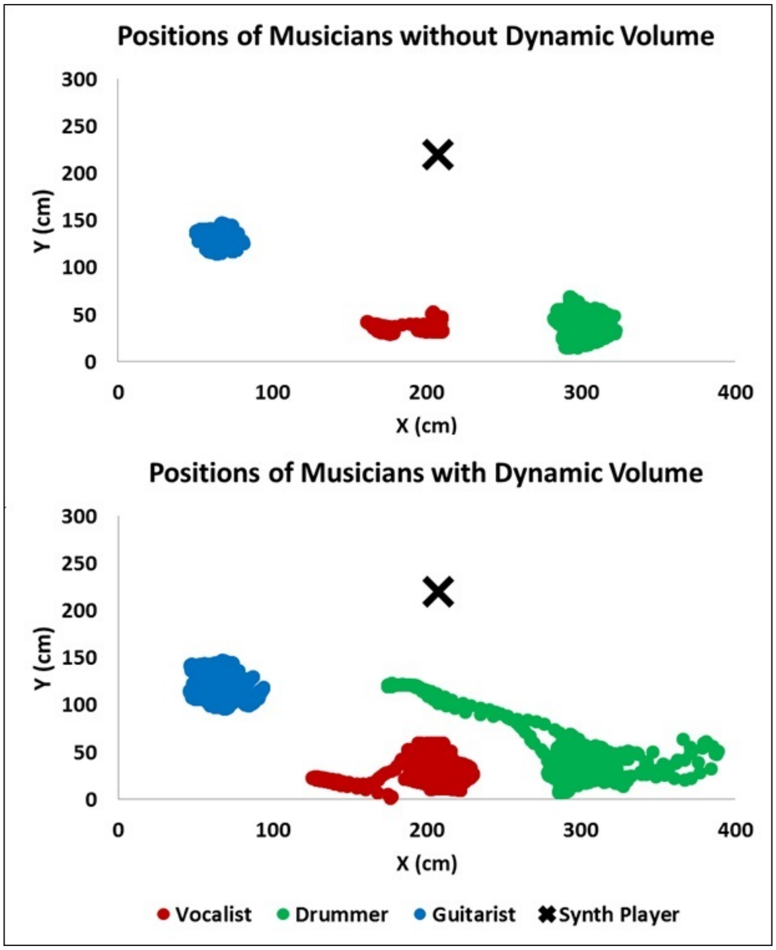

Fig. 5.4: Positions of rhythm guitarist Band 1 members without and with dynamic volume.

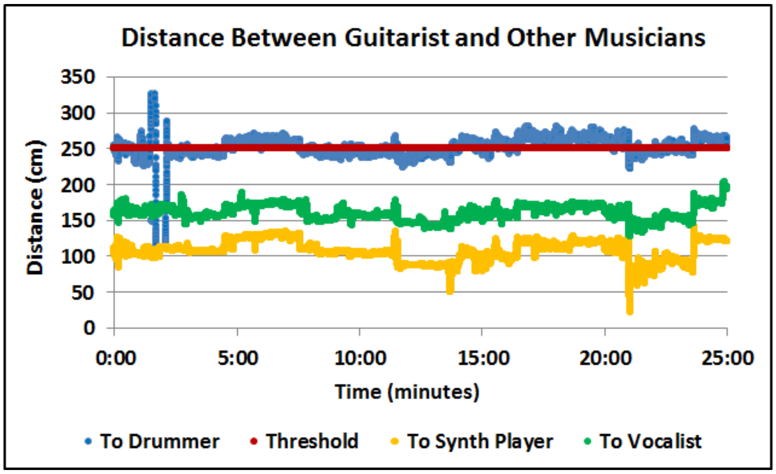

Fig. 5.5: Distances between guitarist and the other musicians in Band 1 when dynamic volume was in use.

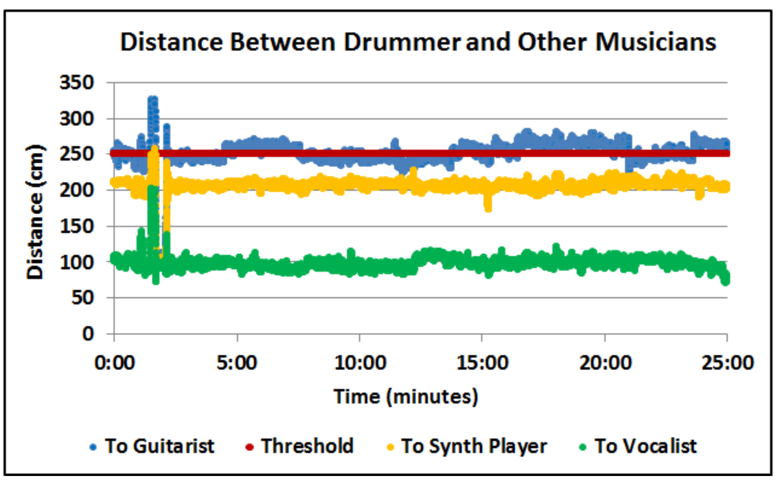

Fig. 5.6: Distances between drummer and the other musicians in Band 1 when dynamic volume was in use.

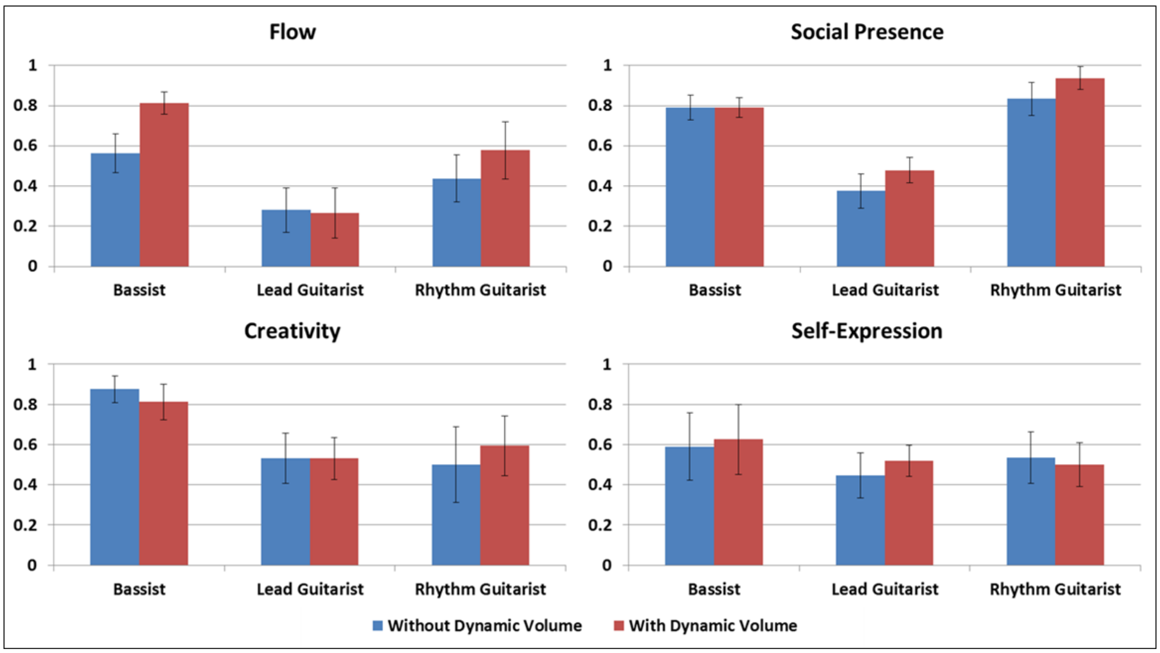

Fig. 5.7: Post-condition questionnaire results for Band 2 members. Averages are shown with standard deviations.

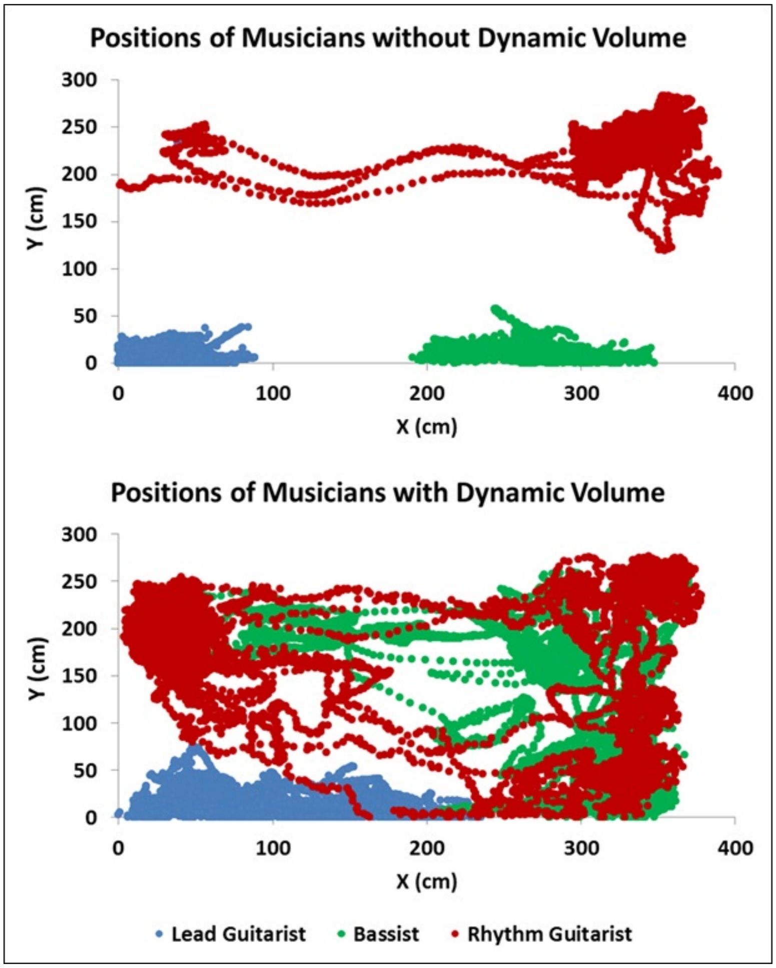

Fig. 5.8: Positions of Band 2 members without and with dynamic volume.

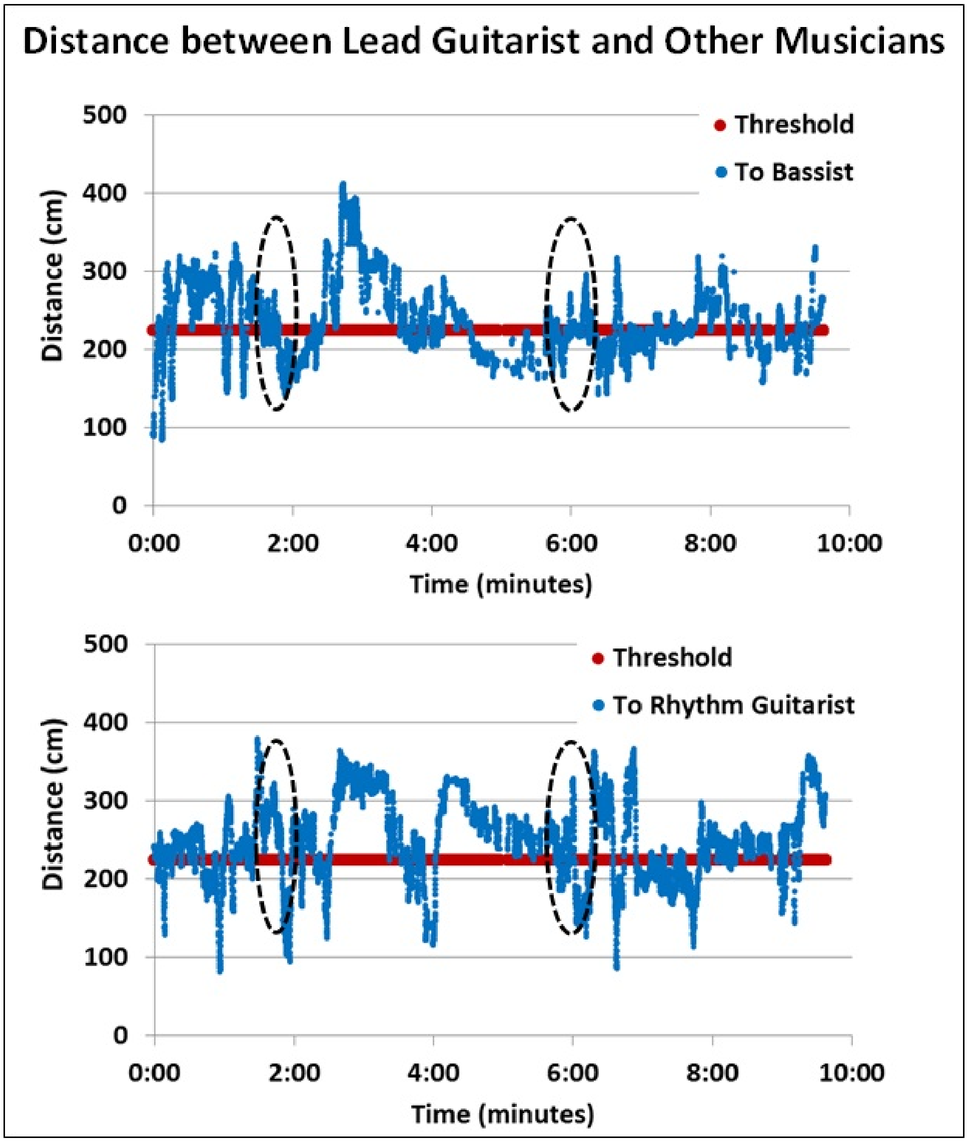

Fig. 5.9: Distance between the lead guitarist and the other musicians over a 10 minute period, or approximately two songs. Two instances of solos have been circled.

Chapter 6

Alpha System

Having verified our first feature in a co-present setting, we set about developing our alpha system by distributing the system across three separate locations In order for the dynamic volume feature to remain valid, however, we had to maintain the illusion that musicians were moving closer and further away from each other, in spite of their remoteness. Therefore, we decided to employ a "shared space" metaphor, whereby each of the musicians' local spaces are mapped onto the Cartesian plane such that they border one another without overlapping, creating, in essence, one large seamless area, as seen in Figure 6.1. This solution, in turn, allows the positions of remote musicians to appear as though located within an extension of each local musician's space. We will refer to those as *virtual locations*. When applied to a scenario with three musicians, the virtual locations of remote collaborators places each of them on either side of the local musician. To maintain this illusion, every location was equipped with two monitors, each displaying a view of one of the remote spaces. To prevent users from falling out of one another's views as they move about their local spaces, a camera was mounted behind each monitor, thereby maintaining a reasonable line of view between the distributed musicians.

As our choice to provide musicians with seamless volume control was met with some success, we contemplated whether augmenting this control with panning, or the ability to control the spread of audio signals between available channels, might be the next natural step in our system's evolution. Thus, our distributed alpha

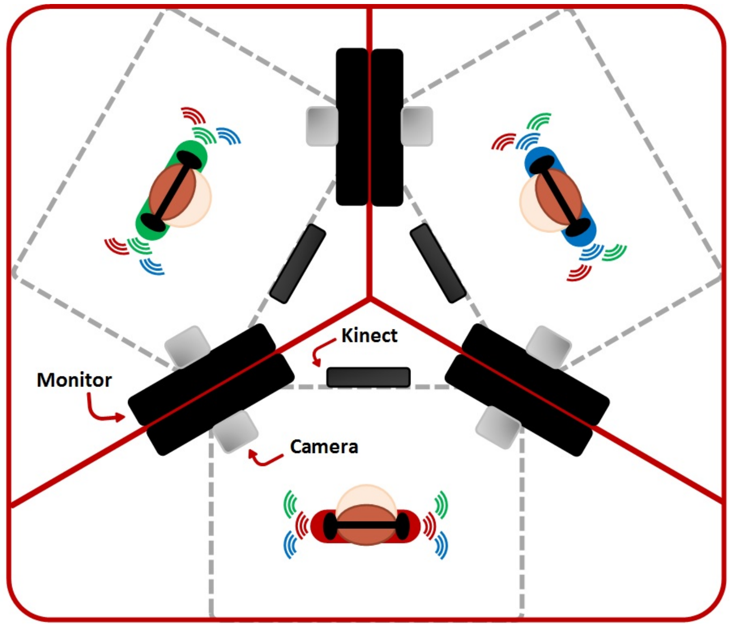

Fig. 6.1: Mapping of three musician locations to create a sense of shared space.

system introduced, along with a fine-tuned version of dynamic volume, the "track panning" feature. Using the latter, a local performer can pan between each of the tracks of the remote musicians simply by tilting his head. The head tilting gesture was specifically chosen due to its explicit nature, in contrast with the seamlessness of changing relative position to control dynamic volume. By employing two different styles of hands-free gestures, one already an integrated aspect of musical performance, and another distinctly introduced to musical performance by our system, we hoped to explore the difference between both styles of interaction. Unlike that behind dynamic volume, and in contrast with the tenets of user-centered design, track panning did not arise as a result of direct user observation, the consequences of which are discussed later in this chapter.

6.1 System Configuration

Our alpha system was deployed across three locations. Although, technically, more than one musician could participate at each location, we chose to focus on the simpler scenario of one musician per location for the purposes of testing and prototyping. The hardware configuration of each space can be seen in Figure 6.2.

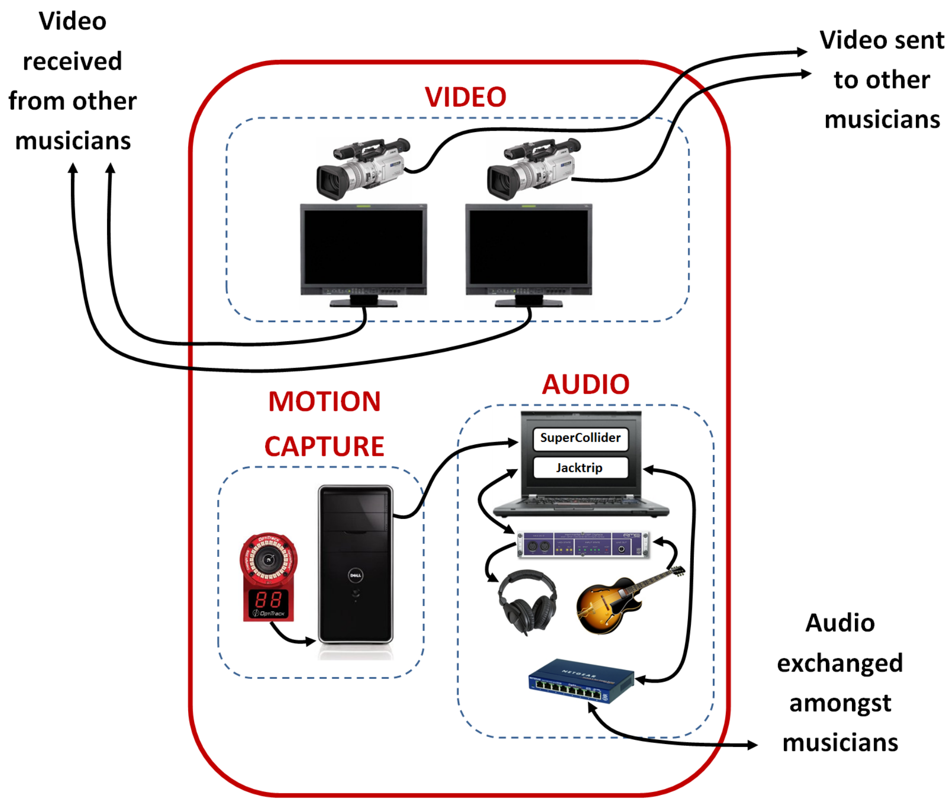

Fig. 6.2: System configuration for our alpha system.

Video

Our experience with the Videoconferencing Privacy System indicated that, while sharing stable, low-latency audio across three machines was relatively easy (using Pure Data's `~nstream` object in that case), video proved to be far more problematic.

In fact, the bandwidth required to operate McGill's Ultravideoconferencing transport proved to be excessive for our Fast Ethernet (100 Mbps) connection, resulting in dropped packets and an unreliable video stream. Given the need for an alternative approach, we opted to create a simple yet stable setup using analog cameras connected directly to production monitors. As described earlier, each location included two monitors, with a camera mounted behind each to maintain a reasonable line of sight across the distributed musicians.

Audio

All streams were processed through JACK[1], a sound server daemon that provides real-time, low latency connections between our audio interface, the Roland Edirol FA-101, and the various software applications listed here. Given our earlier success with SuperCollider, it continued to be our choice of environment for implementing our system features. Audio streams were then shared among all three locations through Jacktrip, a tool for high quality audio distribution developed as part of the SoundWIRE project (described earlier in Section 2.2.2). To further reduce delay and guarantee audio stability, a real-time kernel was used on all machines executing Jacktrip, and a local area network (LAN) was created to connect them through a Netgear ProSafe 8 gigabit switch.

Motion Capture

One of the locations was equipped with a Vicon motion capture system, while the remaining two were fitted with the more portable Optitrack system. Markers were attached to hats worn by the musicians to track their head orientation, while body orientation and position was determined through markers attached to an adjustable elastic band strapped across their chests. Such a configuration provided all the information needed to implement both the dynamic volume and track panning features. Examples of musicians simultaneously using our system at all three locations can be seen in Figure 6.3.

[1] http://jackaudio.org/

Fig. 6.3: Musicians simultaneously using our alpha system, shown at all three locations.

6.2 Latency

We measured the end-to-end latency between locations resulting from our hardware and software configurations using the technique depicted in Figure 6.4. An audio signal was captured through a microphone, then split and routed to two Edirol FA-101 audio interfaces, each being connected to a different computer. Both computers were connected to one another through our Netgear gigabit switch. The first signal was processed via JACK Audio and recorded directly on that computer using Ardour,[2] an open-source audio recording, editing and mixing software on the first computer. The second signal was also processed through JACK Audio, then sent via Jacktrip to the first computer. There, it was recorded via Ardour twice: once after it was immediately received, and again after modification according to our software features in SuperCollider.

An example of the three signals as recorded by Ardour can be seen in Figure 6.5. In total, we recorded 20 sets of such samples, and measured the latency between the signals comprising each set. The average latency introduced when streaming the signal from one machine to another via Jacktrip was determined to be 4.046 ms ($SD = 0.28$ ms), while that introduced by our system features in SuperCollider was an average of 0.376 ms ($SD = 0.087$ ms). We note that when processing audio signals, JACK Audio introduces a constant latency that is dependent on several user-selected parameters, such as the sample rate, the number of frames per period and the number of periods per buffer. While the choice of parameters can, in theory, produce latency levels as low as 0.167 ms, the minimum figure that proved stable enough for our application was 11.6 ms (resulting from a sample rate of 44100 kHz, 256 frames per period and 2 periods per buffer). As a result, the final end-to-end latency between two locations was determined to be, on average, 16.022 ms ($SD = 0.269$ ms), a figure reasonably lower than the ensemble performance threshold, which, as discussed previously in Section 2.2, is typically established to be around 25 ms.

[2] http://ardour.org/

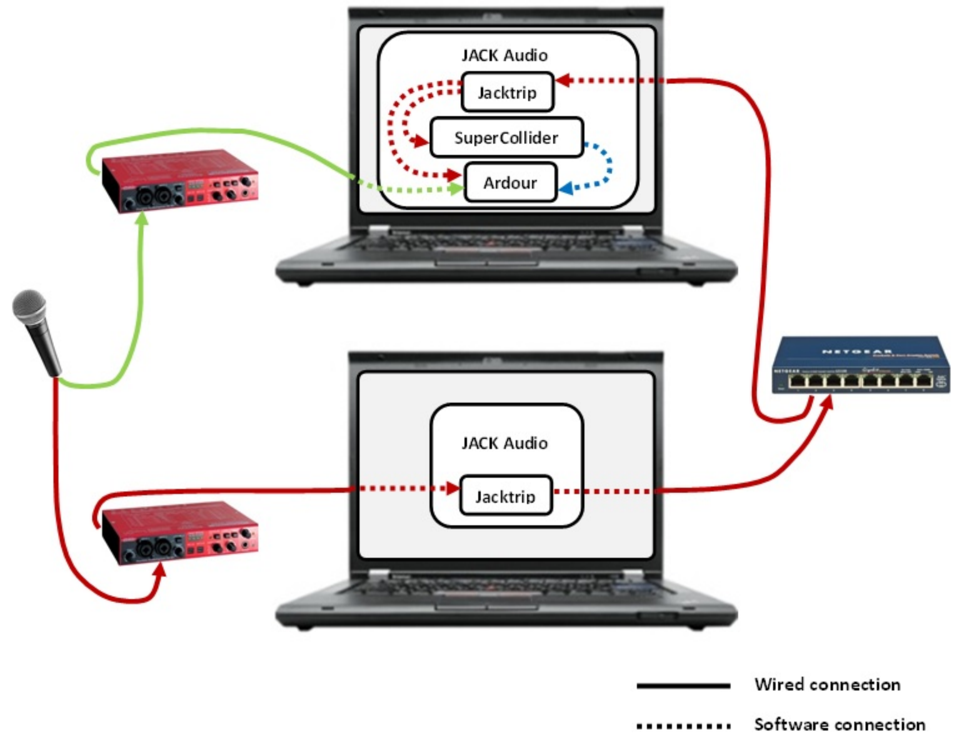

Fig. 6.4: System configuration used for measuring audio latency. The various connection colours represent the three different signals recorded by Ardour for comparison.

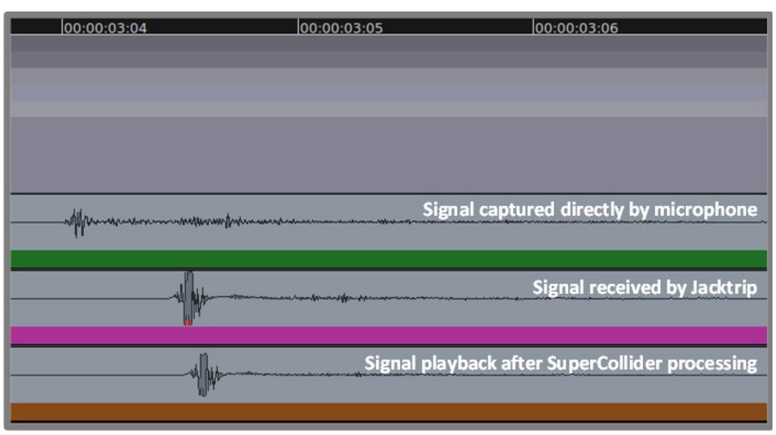

Fig. 6.5: Example of audio signals captured through Ardour to measure latency.

6.3 Graphical Animations

As noted by other researchers, and detailed earlier in Section 2.2.3, musicians typically appeared to make little use of shared video during distributed performance [167]. Furthermore, video alone cannot convey to musicians an understanding of their relationships with one another in a distributed yet "shared" performance space. To remedy this, we designed a graphical representation of the performers, intended to evoke the sense of shared space, by showing where they stood in relation to one another, in spite of their remoteness. In addition, while the effects of auditory features could naturally be heard, we also wanted to provide graphical feedback to further reinforce each participant's state at any given point. Therefore, the alpha system also introduced our first iteration of a GUI that featured animated representations of the musicians. Since we did not want to impose unnatural requirements on the musicians that they focus on a computer screen for any significant amount of time mid-performance, a central objective of this display was that all the information it conveyed could be easily understandable in a matter of seconds. As a result, we began by conducting a preliminary user experiment where we polled subjects on simple graphical representations of moving musicians, as well as volume and panning levels, and used the results to drive our design.

6.3.1 Preliminary User Experiment

During the experiment, subjects were presented with simple example scenarios of musicians performing with one another, and experiencing either dynamic volume or track panning. They were then asked to draw a top down view that best represented each scenario, and subsequently explain their drawing in detail to the test facilitator. Subjects were explicitly asked to draw musicians from a top-down view, along with any volume or panning information provided with each scenario. The test facilitator made sure to emphasize that the drawings were to be kept as simple as possible throughout. In order to ensure the universality, and in turn clarity, of any graphical representation methods we might choose as a result of the outcome of our experiment, we chose not to limit our participants to musicians only, but opted instead to work with a wider variety of subjects. Therefore, out of the seventeen subjects who participated in our study, only 9 participants considered themselves musicians. Of the total number of subjects, 11 were female, 5 were male and one preferred not to answer. Their ages ranged between 19 and 40 years.

The results of the study are shown in Figures 6.6 to 6.8. Sectors of the pie charts labelled as "Others" refer to the total number of suggestions that were named only by one participant. As seen in Figure 6.6, the majority of subjects suggested that the musicians, as seen from a top down view, should be represented by circular markers. Naturally, position information is implicit in such a representation. Figure 6.7, on the other hand, illustrates that just over half of the participants believed changes in volume could best be illustrated through "waves" emanating from the markers, and varying in number in accordance with the sound levels. Unfortunately, as seen in Figure 6.8, there was no consensus when it came to the graphical depiction of panning levels. As a result, we were left to design such a representation ourselves. Based on suggestions from colleagues who were also musicians, we decided to draw the musicians' bodies, seen as "shoulders" from above, in addition to their heads, which would continue to be represented by the circular markers suggested by our experiment subjects. Panning levels would then be illustrated through the head marker sliding towards either side of the shoulders, almost representing a crossfader, a very common audio mixing tool. Such a representation has the advantage of also allowing us to provide body orientation information, which circular markers do not. Nonetheless, since it was not directly suggested through our preliminary experiment, we hoped that any lack of clarity of the metaphor would be uncovered in subsequent user tests with our responsive environment.

6.4 User Controls

In addition to invisibility, Cooperstock et al. list two other factors as critical to the usability of reactive environments: feedback and manual override [51]. Therefore, while we needed a GUI to house the visual representations we had designed, we also wanted to provide users with complete control over the system, and continuous

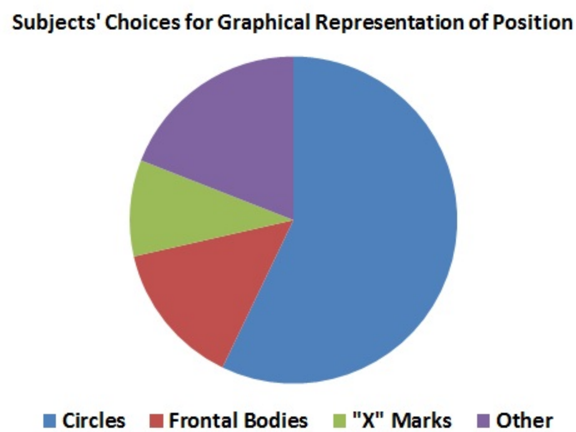

Fig. 6.6: Suggestions made by subjects for graphical representation of musicians.

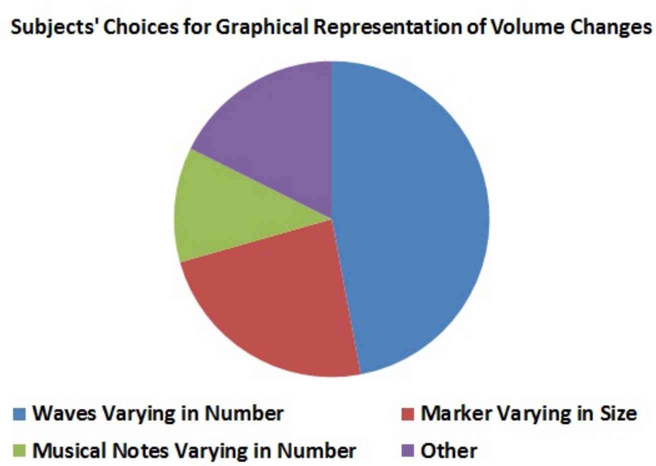

Fig. 6.7: Suggestions made by subjects for graphical representation of volume changes.

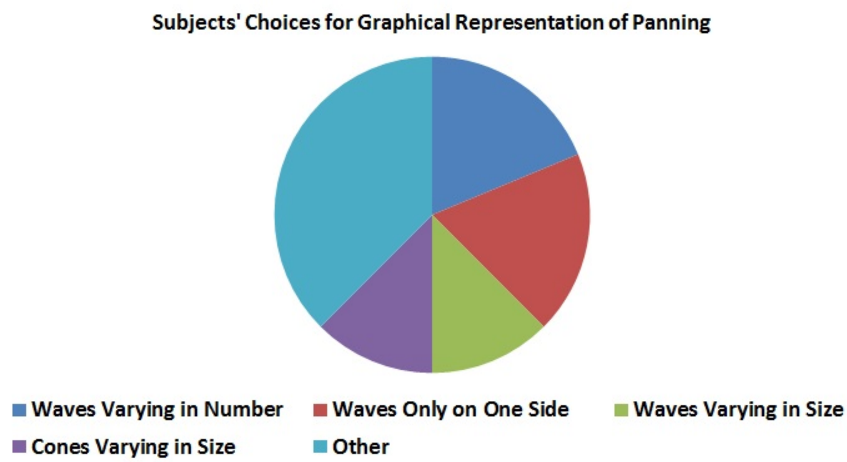

Fig. 6.8: Suggestions made by subjects for graphical representation of panning levels.

feedback over its state. Our goal was to provide even non-technical users the ability to operate and understand our system entirely on their own, without relying on the help of experts to carry out what should, in theory, be simple operations such as setting connections, starting and stopping the system, or changing preferences.

When users first launch our SuperCollider software, they are presented with two introductory screens that allow them to configure the system and set parameters for their jam session. The first introductory GUI, seen in Figure 6.9, allows musicians to choose the system features they would like to use, such as dynamic volume or track panning. After clicking the 'Next' button, users move on to the second introductory GUI, seen in Figure 6.10. Musicians are each assigned a number at the start of a session. When that number is selected from the drop-down menu, the IP addresses of the remote musicians (based on their connection to our gigabit switch) are automatically filled in to prevent errors. One can override this feature by manually entering other IP addresses. When ready, the user then clicks the 'Next' button to move on to the main GUI seen in Figure 6.11.

The main GUI provided participants with the following features:

- **Connecting with Others:** Once a musician clicks the 'Connect' button,

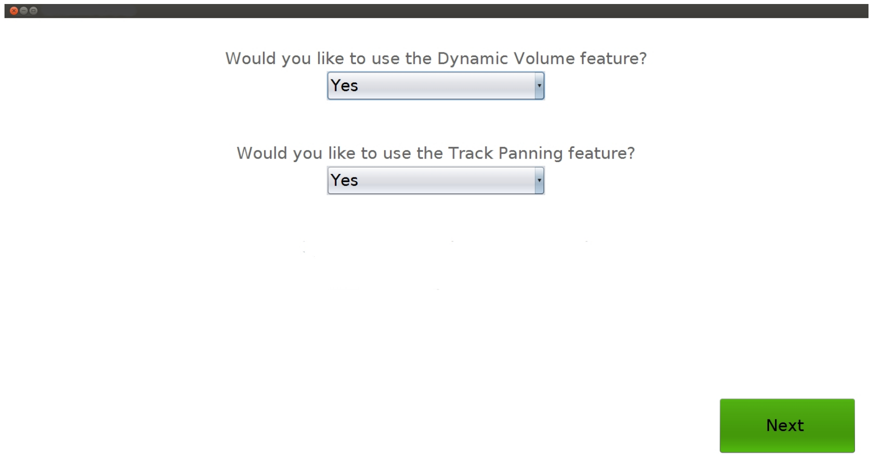

Fig. 6.9: Alpha system GUI, first introductory screen. Users can choose the system feature they would like, and whether to log data.

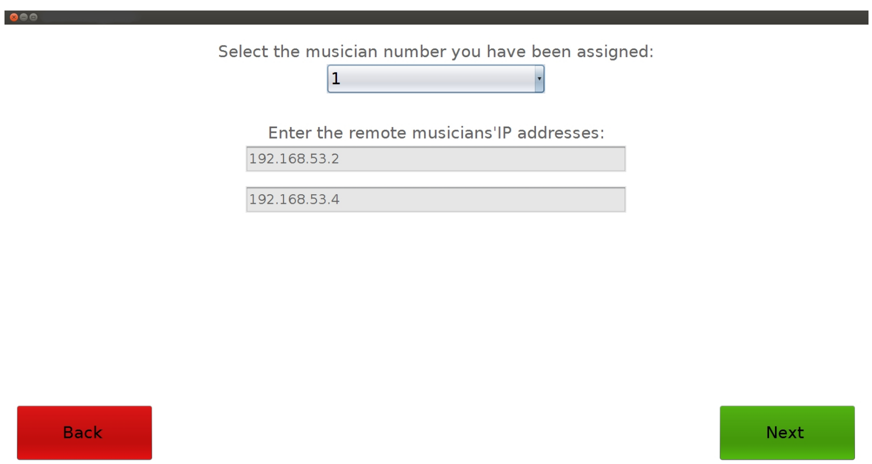

Fig. 6.10: Alpha system GUI, second introductory screen. Users can choose the default IP addresses or change them.

we consider him to be “online”. He is told whether any other musicians are also online, and once all three participants are connected, the ‘Start’ button is enabled to indicate that they can now begin their session, as depicted in Figure 6.11.

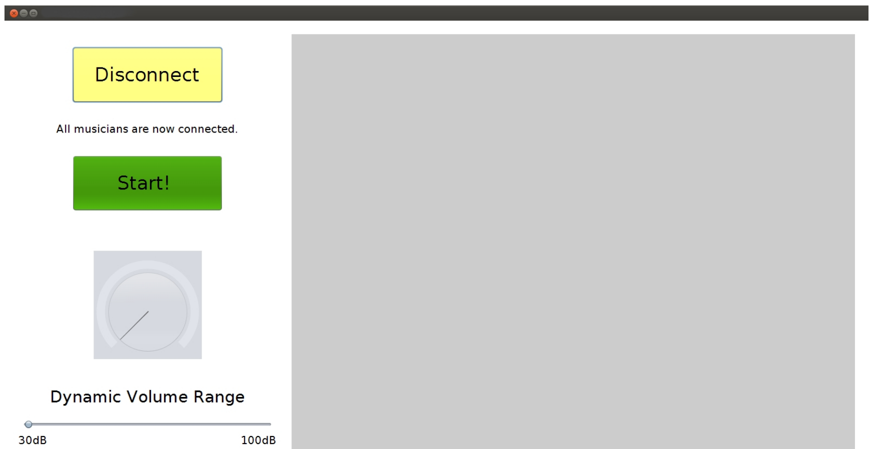

Fig. 6.11: Alpha system GUI, main screen. Both remote musicians are now connected. The session can now begin, and the ‘Start’ button is now enabled.

- **Starting and stopping the performance:** As each musician clicks the ‘Start’ button, he becomes able to hear through his headphones any other musicians who have also pressed their corresponding ‘Start’ button. As seen in Figure 6.12, the ‘Start’ button also turns into a ’Stop’ button for later use.

- **Dynamic visual information:** Once a session begins, musicians are presented with dynamic graphical representations of one another’s positions, orientation, volume and panning level, in accordance with the results of our preliminary experiment detailed above. A closeup of such animations can be seen in Figure 6.13.

- **Adjusting Settings:** Each musician can set his own volume using the circular knob on his GUI. This level is the base volume heard by himself and his bandmates, and subject to increase when dynamic volume is in use. The GUI

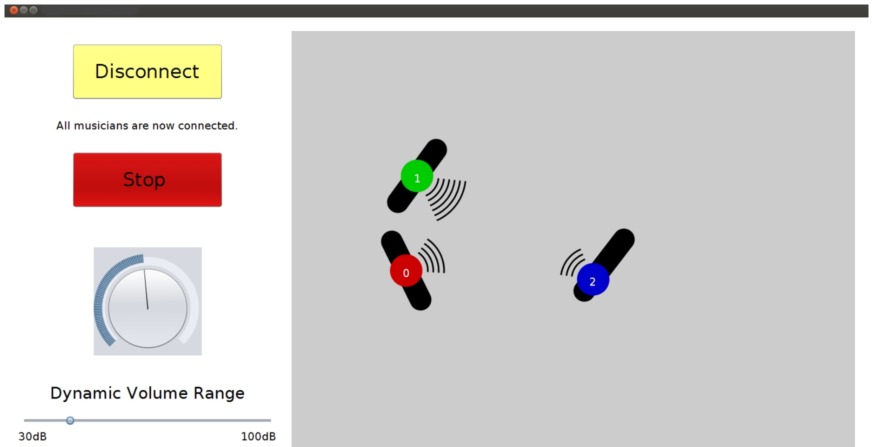

Fig. 6.12: Alpha system GUI, main screen. Performance is underway, with animated avatars representing the musicians as seen from a bird's eye view.

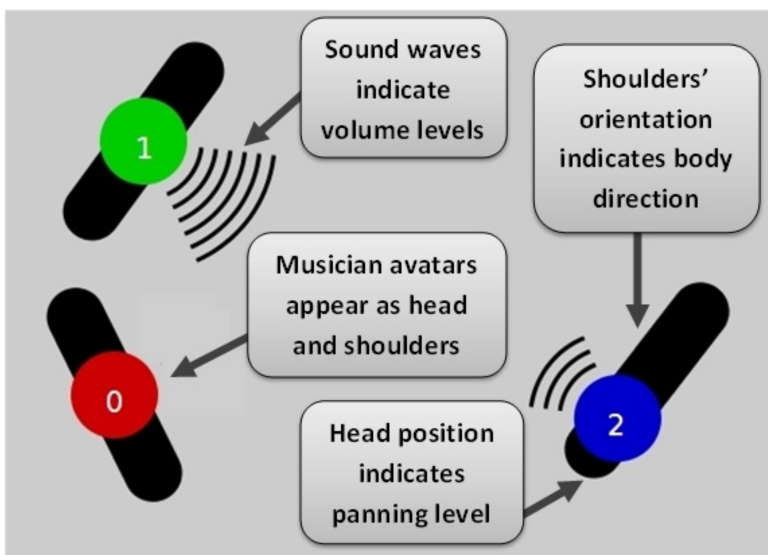

Fig. 6.13: A closeup of the graphical animations used as part of the alpha system's GUI, as seen from Musician 0's perspective (i.e., on his computer monitor).

also includes a slider that allows each musician to independently adjust the sensitivity of the dynamic volume feature, with higher sensitivity leading to greater volume increases over shorter distances travelled.

6.5 Shared Video vs. Graphical Animations Experiment

As described earlier, an additional incentive for adopting a non-video graphical representation was to investigate whether such animations could serve as an alternative to the bandwidth-demanding shared video typically used as the only form of visual communication in distributed performance. Doing so could help make distributed performance accessible to a greater number of musicians, who, under normal circumstances, do not experience the ideal conditions afforded by a gigabit switch and directly wired camera-to-monitor connections. Thus, inspired by other research exploring new visualization techniques in distributed performance, as previously detailed in Section 2.2.3, we used our system to conduct an experiment examining the benefits of front-facing shared video vs. graphical animations on distributed musical performance.

6.5.1 Methodology

For this experiment, members of participating three-piece bands that fit out “Rocker” personas performed together using our responsive environment, while being exposed to one of four “visual conditions” covering all possible combinations of video and graphical animations: No Video or Animations (NVA), Video Only (VO), Animations Only (AO) or Video and Animations (VA). Subjects were exposed to each condition for approximately 20 minutes (3-4 songs), after which they were asked to complete a post-condition questionnaire. The conditions were presented in random order, and all members of a given band experienced the same condition simultaneously. At the end of the experiment, the musicians filled out a final post-test questionnaire, where they ranked the visual conditions in order of preference. The experiment was conducted with three bands, for a total of nine musicians, all male, and ranging from in age from 19 to 26 years. Members of each group were required to be familiar with each other, having previously performed together regularly for at least one year.

6.5.2 Results

The post-condition questionnaires polled musicians on their perception of several aspects of the performance, such as their overall understanding of distributed performance, their understanding of the system's features, their ability to learn how to use the system, their awareness of their remote colleagues, their comfort levels and the likelihood they would perform under the same conditions again in the future. As seen in Figure 6.14, all categories, with the exception of "Ease of Learning", received the highest scores under the Video and Animations condition. The results for "Ease of Learning" are consistent with our expectations, however, since interpreting avatars, no matter how simple in design, will naturally introduce additional learning time. In addition, analysis of variance indicated that the visual conditions had a significant main effect on the musicians' "Understanding of distributed performance", with $F(3,24)=4.983$, $p=0.0079$, and "Likelihood to use again", with $F(3,24)=5.34$, $p=0.0058$.

According to the post-test questionnaire, Video and Animations was ranked as the top preference by six out of the nine test subjects. The remaining three chose Video Only instead. Additional comments elicited in the questionnaire indicated that performance without video was rather difficult, leading to a breakdown in non-verbal communication that, in turn, contributed to a lack of awareness between band members. Furthermore, some of the musicians explained that, while they did not find it necessary to stare directly at their video monitors mid-performance, keeping the video streams in their periphery was crucial in maintaining an awareness of their remote colleagues. Thus, while our observations of the musicians' interaction with their video monitors were consistent with other research findings (cf. the work of Schroeder et al. [167]), we also learned that it would be highly detrimental to use them as justification for eliminating shared video altogether.

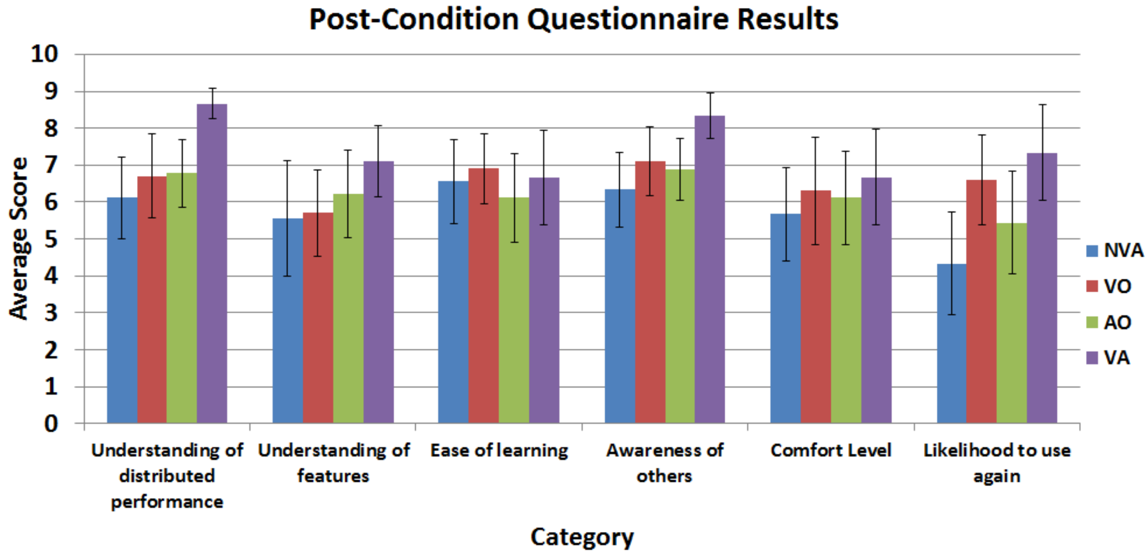

Fig. 6.14: Result of post-condition questionnaires during Video vs. Animations experiment. Averages shown with standard deviations.

6.5.3 User Feedback: The Net-Music 2013 Symposium

The same trio of rock musicians consisting of a lead singer and rhythm guitarist, a lead guitarist, and a bassist, who had participated in testing an early prototype as described in Section 5.4.2, performed using our alpha system as part of a demo for the "Net-Music 2013: The Internet as Creative Resource in Music" symposium. Overall, the musicians required very little training to become familiar with the system and its functions. When asked if the latency was perceptible, they reported not finding it at all problematic. They were easily able to find suitable volume levels, and dynamically adjusted them throughout the performance through the dynamic volume feature. In comparison, however, track panning was used far less often. The bassist explained that he did not quite understand the use for track panning, as dynamic volume seemed to provide him with enough volume control. He added, however, that he could perhaps grasp its usefulness after using the system over more extended periods of time. This is perhaps not entirely surprising, given that, unlike dynamic volume, which evolved as a result of direct user observations, track panning arose from our own assumptions about the potential benefits of providing musicians with panning capabilities. Given that such an approach does not directly align with the mandates of standard user-centered design, track panning would prove more difficult a feature to adapt to the musicians' expectations, as further described in the following chapter.

Finally, the musicians commented on the GUI several times, especially with regards to the visual representations, often joking when their avatars were getting increasingly closer to one another as they moved about, even though the musicians were obviously in different rooms. This helped indicate their perceived sense of shared space. However, when asked about the usefulness of the avatar representation, the bassist explained that he was unsure of its purpose. He described the variations in volume triggered by the dynamic volume function as being a "good feedback for distance" and, therefore, he did not find it necessary to gauge that information from the avatars themselves. He added, however, that with time, he might find the visual representations more useful. The lead guitarist reported finding the visual layout of the GUI, especially the volume knob and dynamic volume sensitivity slider, to be "very simple to use and very responsive". Overall, the musicians found our performance environment enjoyable and easy to use, with the lead guitarist adding that he "saw great potential in the arrangement".

Chapter 7

Beta System

Our Video vs. Animations experiment demonstrated that the visual condition most suitable for our responsive environment is Video and Animations. However, the experiment brought to light another interesting consideration: comments left by a number of musicians in the post-test questionnaires indicated that they found the system to be, above all, quite novel and exciting. In fact, we had received similar feedback from other bands when testing our early prototypes. Therefore, we wondered whether the positive feedback was simply the result of a novelty factor. Furthermore, as described in Section 6.5.3, some musicians had previously explained to us that while the usefulness of the dynamic volume feature was immediately apparent to them, they felt the need to experiment with the track panning feature for longer than a test session before they could form a more accurate impression. Finally, when musicians provided suggestions for the improvement of the system features themselves, we realised that the 'one-off' nature of traditional formal experiments did not provide us with an opportunity to test the effects of small, iterative changes to our system on a regular basis. As a result, we decided that the evolution of our alpha system into its following iteration should be driven by a long-term deployment and testing cycle with a smaller number of users.

7.1 System Configuration

At that point, we replaced the motion capture systems we had been using in the earlier prototypes with more affordable Microsoft Kinect units. This decision was motivated in large part to allow our system to reach a potentially wider audience of users. In addition, it allowed us to reduce the number of computers required to implement our responsive environment, as machines dedicated to the operation of motion capture systems were no longer required. As such, the new hardware configuration can be seen in Figure 7.1. Values captured by the Kinect were sent to our SuperCollider software via OSC messages. Otherwise, our audio and video configurations remained the same as with our alpha system, described earlier in Section 6.1.

Seeing as our hardware and software configurations differed little from those of the alpha system, the baseline latency continued to hover around 16 ms, the same figure calculated earlier in Section 6.2.

7.2 Long-Term Deployment

Inspired by Grudin's views on the importance of long-term system evaluations within CSCW research [85], and the success of such a methodology within the contexts of both remote collaboration [123] and musical performance [55], we were motivated to improve on the depth of feedback we had received from musicians during previous user tests, and elicit results beyond simple novelty effects and initial impressions. Thus, we sought to capitalize on the benefits of quantifiable, repeatable user studies and the depth of feedback inherent to participatory design by merging elements of both methodologies into a long-term testing and collaboration cycle.

Therefore, we organized a series of performance sessions with a band consisting of a 25-year-old guitarist, a 26-year-old keyboardist (both of whom also alternated lead and backup vocals), and a 22-year-old bassist. All three had performed together regularly (approximately once a week) for almost two years. An introductory brainstorming session was first held, allowing us to showcase our system to the band

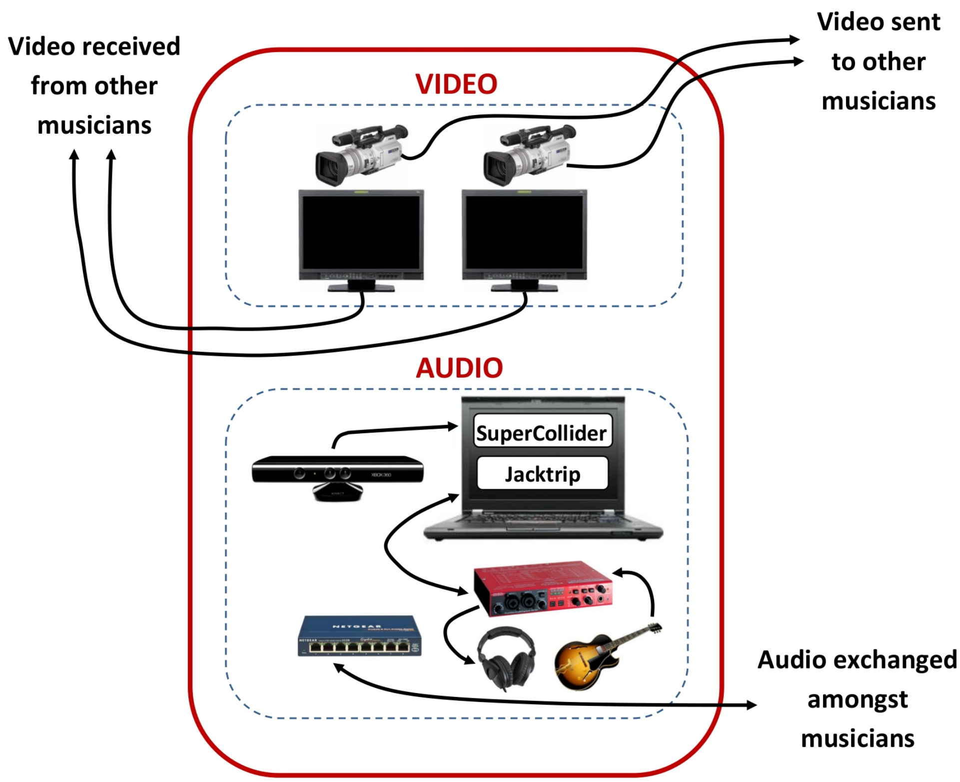

Fig. 7.1: System configuration for our beta system.

members, and discuss our vision for the long-term collaboration. Subsequently, we organized weekly meetings that combined formal, quantitative tests with informal, qualitative discussions. The formal aspect of these sessions was designed to evaluate the effects of the system's features on those criteria of performance that, according to our user interviews (described in Section 5.3) musicians deemed important: interaction with other musicians, enjoyment, creativity and self-expression. Since our goal was to "to discover (rather than to verify)" the effects of each system feature on these various aspects of performance, we knew that the qualitative experiment framework proposed by Ravasio was most suitable to our needs [155]. Therefore, we employed both of her techniques of separation/segmentation and adjection/intensification to design a number of sessions, each focusing on a different feature of the system through an A/B/A-style test, where musicians performed once without the feature, once with the feature, then once again without the feature. At the beginning of each session, musicians were asked to select their base volume and reverb levels collaboratively. It is those base levels that our system features would subsequently affect during condition B. Each condition lasted for approximately 15-20 minutes, or the time it took the musicians to play through three songs. As with our early prototype evaluations, we maintained an experience-based approach by asking musicians to meet an active goal of playing a fixed number of full songs, while expressing any feelings and concerns, rather than carry out any specific tasks under each condition. The musicians also completed the post-condition questionnaires described earlier in Table 5.3, and designed to assess the performance criteria listed above. Position and orientation data were collected throughout, along with video footage and audio recordings. After the formal test component of each session, an open discussion in the style of a non-leading interview was held. Musicians were loosely probed about their approach towards the performance and their feelings about the system, and encouraged to provide criticisms, along with suggestions for improvement.

7.2.1 Session 1: Musician Spatialization

Our first session with the band was designed to focus on a new "musician spatialization" feature, whereby the sounds of remote instruments are perceived as emanating from the correct spatial position within the musician's local environment. This feature was first suggested to us by another musician who had tested the alpha system, as a possible improvement over track panning, the use of which, as explained earlier, was not immediately apparent to some performers. The idea was to mimic the spatialization effects naturally experienced in a co-present setting, where performers can easily perceive the distance and direction of other instruments surrounding them based on their position and orientation.

Unlike the addition of certain other features such as track panning, described next, musician spatialization was intended simply to recreate some of the natural acoustic dynamics that are lost in typical distributed performance. As such, unlike other system features, no explicit gesture is required to activate it: as long as the feature is enabled, the audio from remote musicians will continue to be rendered. However, our post-test discussion with the musicians revealed that the "passive" nature of the feature had left them somewhat confused. The guitarist, for instance, explained:

"I could tell there were changes happening when there were changes happening, but I really had difficulty at times making sense of it."

Although its mapping was discussed with them before the performance, they continued to look for a "triggering" gesture that would allow them to control the effect. Nonetheless, as the musicians reflected on their performance after the feature was once again explained to them, all three of them indicated they would be very inclined to try it again in light of their new understanding. They were eventually given another opportunity to test musician spatialization in Session 5, described below.

Analysis of position and orientation data did not reveal any significant changes in behaviour when musician spatialization was used.

7.2.2 Session 2: Track Panning

The second session focused on the track panning feature, again through the form of an A/B/A test, followed by a discussion. While previous testing had shown that the usefulness of track panning was not immediately clear to some musicians, we opted to test this feature once more in a bid to better understand its problematic aspects, by taking advantage of the more profound level of feedback afforded by the long-term deployment sessions.

As seen in Figure 7.2, orientation data from the formal tests indicates that, while all three musicians experimented with the feature, the keyboardist and guitarist felt more inclined to sustain their interaction with the feature for longer periods of time. The guitarist, in particular, regularly isolated the bass track by turning his head to the left, as it helped him maintain his rhythm. In the post-test discussion, the bassist revealed that since his instrument's low frequency already made it harder to distinguish from the others, he was less keen on disturbing the base volume levels coming through his headphones. This issue of effective mixing was raised again, and subsequently resolved, during the dynamic volume tests described next.

During the post-test discussion, some of the musicians criticized the head-tilting gesture of the track panning feature, noting that it would feel more "natural" to turn one's *body*, rather than tilt one's head, towards the virtual location of another musician on whose track they wanted to focus. The guitarist, for instance stated:

"I found the movement, the motion a bit unnatural... It would make a bit more sense if like, if say I'm looking at [bassist's name] and [keyboardist's name], if I wanna hear more of [keyboardist's name] I can just turn to him. If that was the pan, that would be more truly obvious to me."

Thus, the head tilting gesture, which, as explained earlier, was specifically chosen due to its explicit nature, did not suit the seamlessness the musicians felt should be inherent to panning. Nonetheless, the musicians did appreciate the practical aspect of the function, with the keyboardist explaining:

"Well, mid-performance, say there was a part in the song where a few people were harmonizing together, if I could turn to the screen and we could hear each

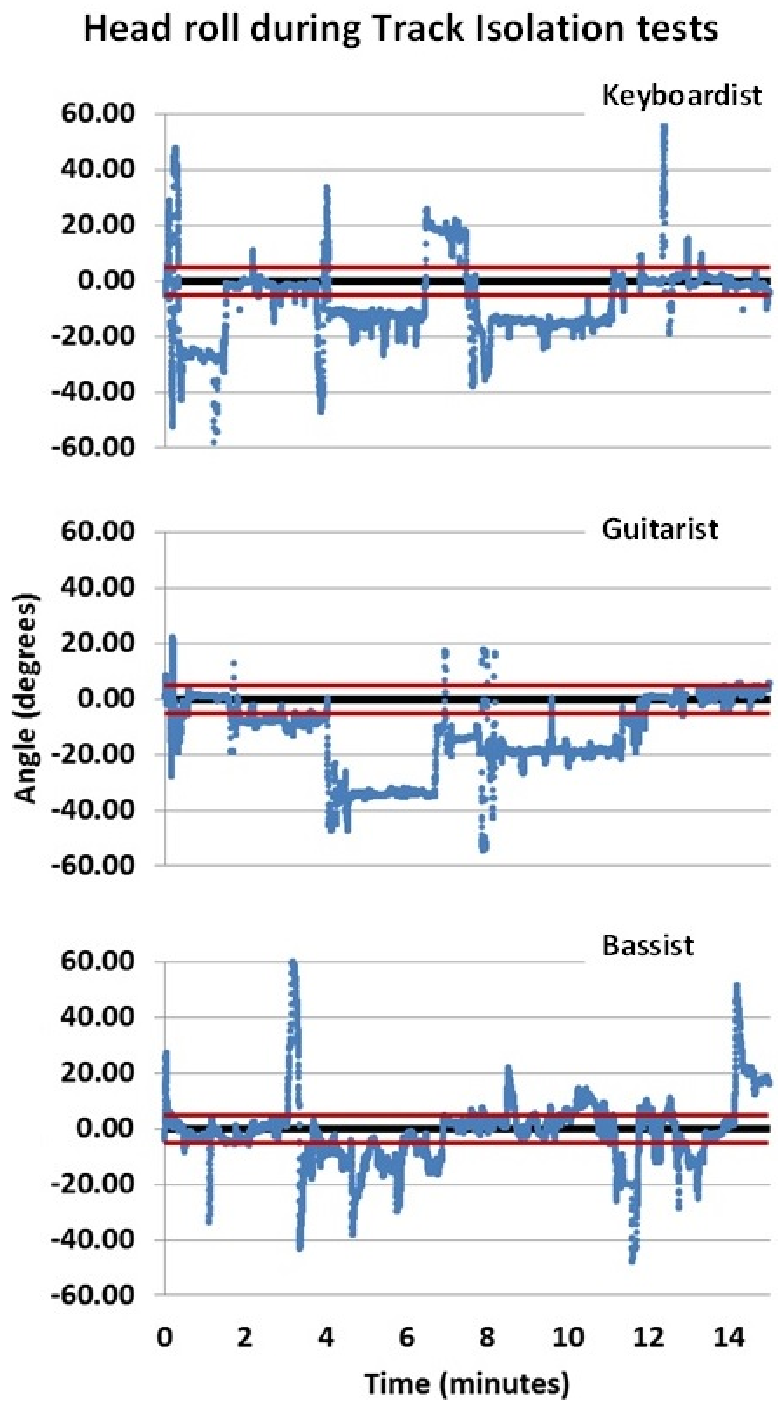

Fig. 7.2: Head roll, or tilting, data for all three musicians when track panning was used. The red line represents the threshold of +/-5 degrees, beyond which the feature was activated.

other better that way, like that would be practical for sure."

The track panning feature was subsequently updated to respond to body direction instead, and musicians had an opportunity to test this modified version in Session 5, as described below.

7.2.3 Session 3: Dynamic Volume

The third session included an A/B/A test of the dynamic volume feature. Analysis of position data, shown in Figure 7.3, revealed that the use of this feature helped encourage all three musicians to increase the range of space they covered, rather than maintaining a fixed location, as they were inclined to do otherwise.

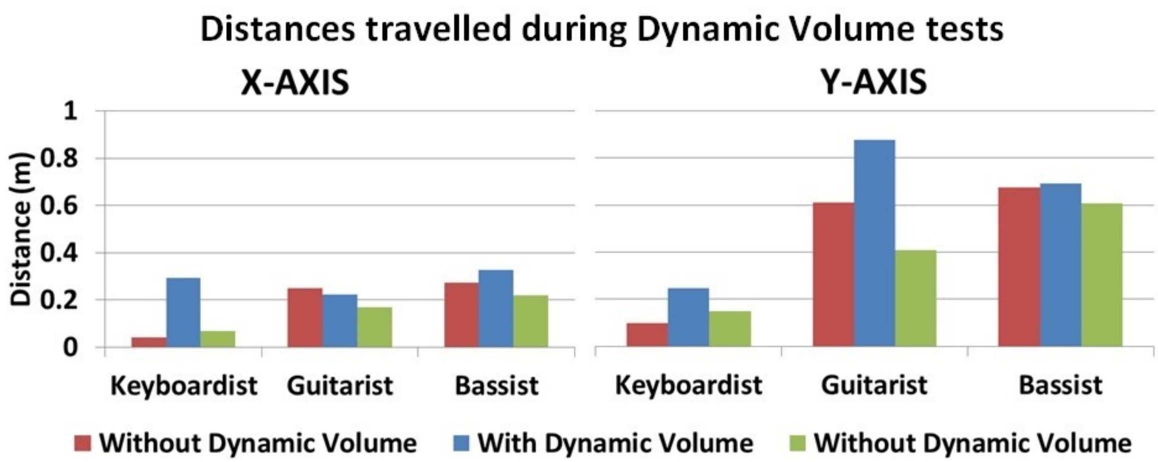

Fig. 7.3: Maximum distances travelled by all musicians during dynamic volume tests.

Up until this point, dynamic volume was implemented such that moving closer to the virtual locations of other musicians would increase the perceived levels of their instruments. During the post-test discussion, however, the bassist explained that he did not feel the need to hear his band mates' levels getting louder, as he found that all members were already setting their volumes as loud as they wanted at the start of the performance. In addition, when his perception of the other musicians' volumes increased as they approached his virtual location, his own instrument became "drowned out" in the overall mix, given the low-frequency nature of the bass.

As a result, we collectively agreed the feature should be altered such that the base levels instead *decreased* as the musicians moved away from each other's virtual locations. While this would still mean that a local musician would perceive his remote colleagues' volumes as getting louder as he approached their virtual locations, it effectively rendered the chosen base volumes as the maximum, rather than minimum, levels experienced. The updated dynamic volume feature was subsequently made available to the musicians in Session 5, described below.

During the post-test discussion, the musicians also began expressing their interest in controlling another aspect of their sound beyond volume level. The keyboardist, for instance, explained:

"I don't know if this is getting too out there, but also for a novelty effect it would be cool to somehow have psychedelic video. It would definitely be amusing."

Although we sought to fully incorporate the musicians' recommendations, we also wanted to steer away from idiosyncratic suggestions. Thus, we explained that perhaps the appeal of psychedelic video could be a matter of taste, and that our goal was to create system features that, along with their potential for artistic expression, held some degree of usefulness to most musicians, rather than sheer novelty. This led to the suggestion of reverberation as a more suitable addition that could enhance the element of creativity, and allow the musicians to experiment with different sounds. According to the musicians, an increase in reverb when moving further away from each other's virtual locations could further help enhance their feeling of shared space, giving them a more concrete sense of dimension due to the effect's "echoing" nature.

7.2.4 Session 4: Dynamic Reverb

Inspired by our discussion with the musicians at the end of the previous session, we began developing the idea for a "dynamic reverb" feature that would allow musicians to experience increasing levels of reverberation as they move further away from one another's virtual locations. We held an interim session during which the musicians were invited to experiment with reverb used to simulate rooms of different sizes, and participate in designing the overall effect. Subsequently, the fourth session was centered on the A/B/A testing of the newly implemented "dynamic reverb" feature. In the post-test discussion, the musicians revealed that they were quite pleased with the feature, with the guitarist stating:

"I felt that it kind of reacted how I would have wanted it to. It felt a bit like I was able to use it and predict how it was gonna be a bit better. It was cool."

Similar to the earlier dynamic volume, dynamic reverb helped increase the interpersonal interaction between musicians, and encouraged them to take full advantage of the available space (see Figure 7.4).

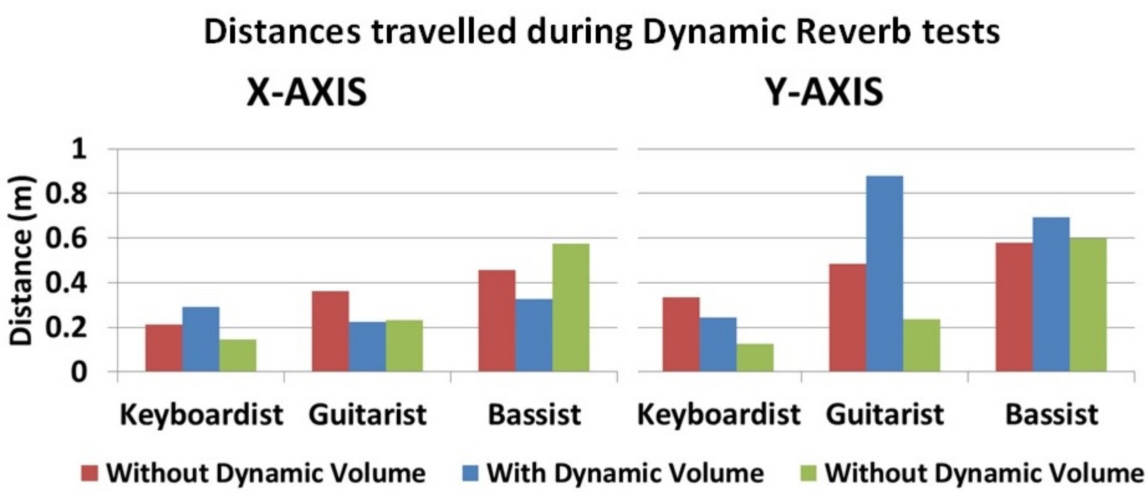

Fig. 7.4: Maximum distances travelled by all musicians during dynamic reverb tests.

7.2.5 Session 5: Freestyle

The fifth session was a "freestyle" performance: the musicians were simply asked to jam for an hour, selecting which features to turn on or off throughout according to their needs. This session also introduced a new feature that emerged from a participatory design cycle we held with a composer, described in full detail in the next chapter. Dubbed "mix control", this new addition allows a local musician to listen to his own instrument being mixed with either of the remote musicians' one at a time, simply by tilting his head in the direction corresponding to the remote musician's virtual location.

The session lasted approximately 50 minutes. Figure 7.5, which illustrates the musicians' usage patterns of our system features, reveals that dynamic volume, dynamic reverb and musician spatialization were the only features that were turned on throughout the entire duration of the session by at least one musician. In contrast, track panning and mix control were only used intermittently. This can, in part, be attributed to the fact that the dynamic volume and dynamic reverb features lend themselves quite well to experimentation: a musician may seamlessly move from one position to another to alter his mix until reaching a "sweet spot", where he may remain for as long as necessary, or until he feels the desire to try another arrangement. Musician spatialization requires no explicit input from users and, as such, is a non-intrusive feature. In addition, all three features exploit natural properties of sound we experience everyday: for instance, increasing the distance between two musicians simulates a larger room with greater reverberation levels; distant sound sources are typically experienced as being quieter than close ones; we perceive sound sources according to their positions around us. As such, these features can be considered less intrusive or distracting than track panning and mix control, which rely on gestures that are more pronounced in nature than simple motion and produce a more complex effect, thereby requiring a greater cognitive load. Furthermore, mix control and track panning were designed to address needs that are comparatively more utilitarian in nature, allowing musicians to listen more attentively to each instrument in the mix and, in turn, make any necessary adjustments to their own sound. An example of this was provided by the keyboardist during the post-test discussion, as he described using track panning to isolate the guitarist so that he could, in turn, figure out what his own accompanying chord progression should be. As a result, while the musicians' perceived relative importance of each of the various features at any given moment is likely a confluence of several factors—including their familiarity with a particular song, the arrangement of the song, and even their moods—we suspect that they found such features as track panning and mix control to be more effective as short-term solutions addressing specific needs.

Figure 7.6 shows the total length of time for which each feature was used by the individual musicians, and points to dynamic volume and dynamic reverb as the

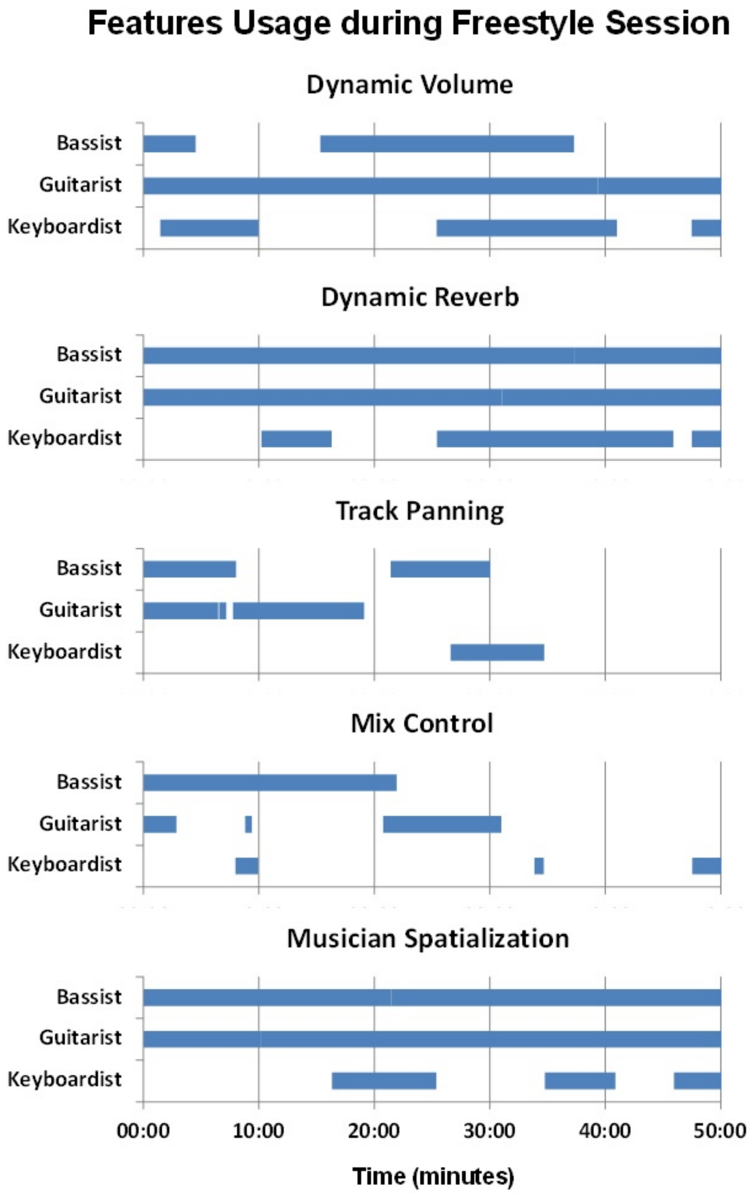

Fig. 7.5: Usage pattern of features by the musicians during freestyle session.

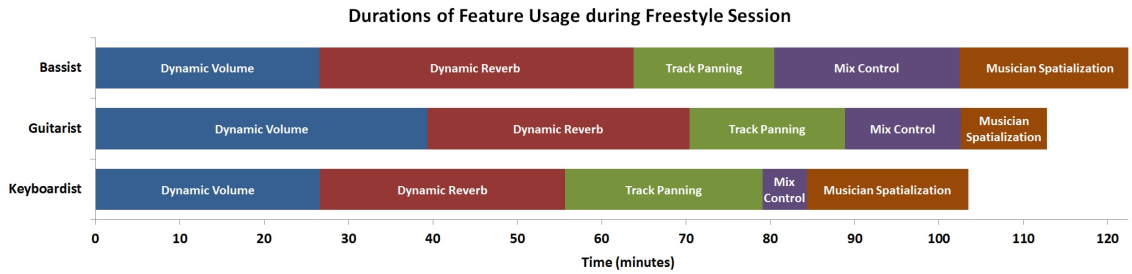

Fig. 7.6: Total length of time for which each feature was used by the musicians during freestyle session.

features most favoured by all three musicians, if only on the basis of duration of use. This was also confirmed during the post-test discussion, where the guitarist and bassist listed the dynamic volume and dynamic reverb features as their favourite. The keyboardist, on the other hand, found mix control to be most useful, as his seated position made active control of the former two features rather difficult.

The post-test discussion also allowed the musicians to provide their opinions of the overall state our system had reached thus far as a result of our on-going collaboration. Feedback indicated that the new version of dynamic volume met the musicians' expectations, and was a great success. More specifically, the musicians explained that if someone's instrument seemed too loud at any point, simply taking a step back would effortlessly help them adjust the mix. In addition, this session provided the musicians with an opportunity to re-visit the musician spatialization feature in light of their improved understanding of its functionality. The latter proved to be a success with the guitarist and keyboardist, who were able to finely control it, now that the mapping had been made clearer to them. When asked whether they would use the system in a scenario where they could not be physically co-located, all three agreed that the features would be quite beneficial in facilitating distributed collaboration. The keyboardist, for instance, stated:

"I think it's like, if we're doing something like jamming in different cities, any sort of software that has extras like that, would be fun... it could be a means to prolong your jam if it's getting boring or something. You could try different sounds or just mess around with it. But there's a practicality to the features too."

Throughout all sessions, the musicians had also been providing feedback on improving the overall sound of the system, recommending preferred volume and reverb levels, and suggesting means to reduce any distortion. By Session 5, all of them were very pleased with how far the system had evolved, describing the sound as far "smoother" and more pleasant to the ear than it was at the start of our collaboration. For instance, when asked how they would gauge the changes in sound quality based on their previous suggestions, the guitarist explained:

"It's definitely come a long way in terms of the quality of the sound that's coming through my ears. So that's the idea, I guess. It sounds good, so that's good."

7.2.6 Additional Aggregated Results

Questionnaire Analysis

As noted earlier, musicians also completed post-condition questionnaires during each of the A/B/A tests. The questions were designed to assess a number of factors, such as the musicians' perceived sense of enjoyment, creativity and self-expression. Responses were tabulated and analyzed to determine the number of musicians for whom each of the system's features helped improve the factors listed above. As seen in Figure 7.7, all features helped contribute to increased levels of enjoyment, with musician spatialization and dynamic reverb performing best in that regard. Furthermore, track panning contributed to an improvement in the musicians' sense of self-expression. Overall, however, creativity appeared to be the factor that benefited least from our system features, increasing only when dynamic volume was in use.

Discussion Analysis

All of our post-test discussions with the musicians were recorded and transcribed before a Qualitative Data Analysis (QDA) was performed. During a repeated coding process, comments were labelled and grouped, until three major categories emerged: Interaction, Sound Quality and Perceived Usefulness. Table 7.1 provides a list of the specific codes that define each category. Comments were subsequently tagged as

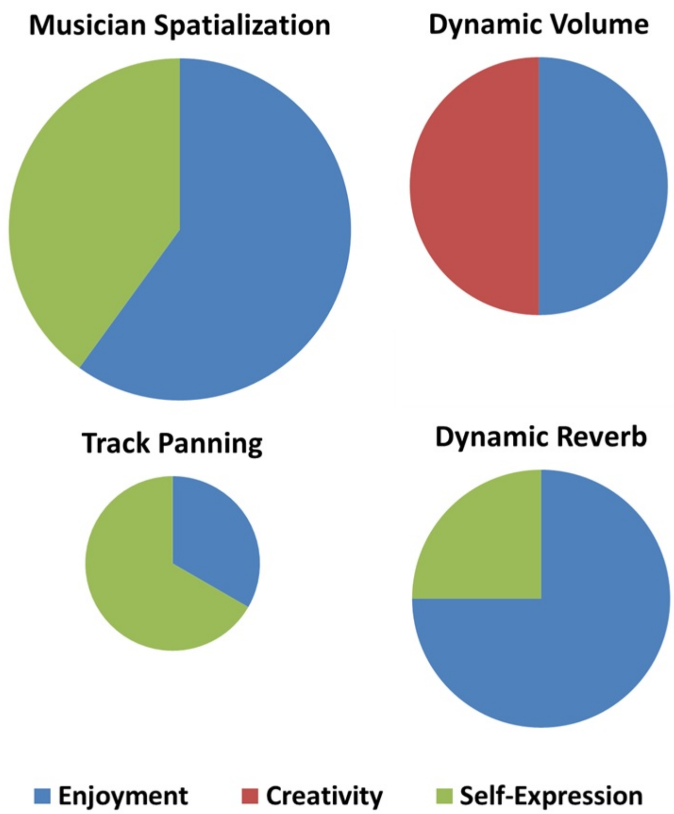

Fig. 7.7: Effect of each feature on performance criteria. The size of each pie represents the total number of times, across all criteria, that an improvement was marked.

"positive" or "negative", and the total count for each per category was tabulated for the sessions.

Category	**Components**
Interaction	Gestures used Control over the feature Sensitivity of the controls Mapping from gesture to resulting sound Clarity of mapping
Sound Quality	Instrument volume Levels Vocals volume levels Clipping and distortion Overall sound
Perceived Usefulness	Likelihood to use again Practicality of feature Envisioned use case scenario

Table 7.1: Descriptions of major categories that emerged during Qualitative Data Analysis of post-test discussions

As seen in Figure 7.8, the number of positive comments for each category slowly improved throughout the sessions, with a particularly sharp increase in the Interaction and Perceived Usefulness categories seen during Session 5. We believe this to be, in large part, due to the nature of the session itself, as musicians were given the opportunity to try out the system features after all the feedback and suggestions they provided had been incorporated. In contrast, Figure 7.8 also shows a steady decrease in the number of negative comments made for all three categories. Together, these result indicate that we were successful in systematically incorporating musician feedback into our system design. In the end, the band members found the beta system that evolved from our weekly sessions to be a vast improvement over its predecessor.

7.3 Latency Experiment

Up until this point, the latency levels achieved through our software and hardware configurations proved to be a non-issue for musicians. This allowed us to focus

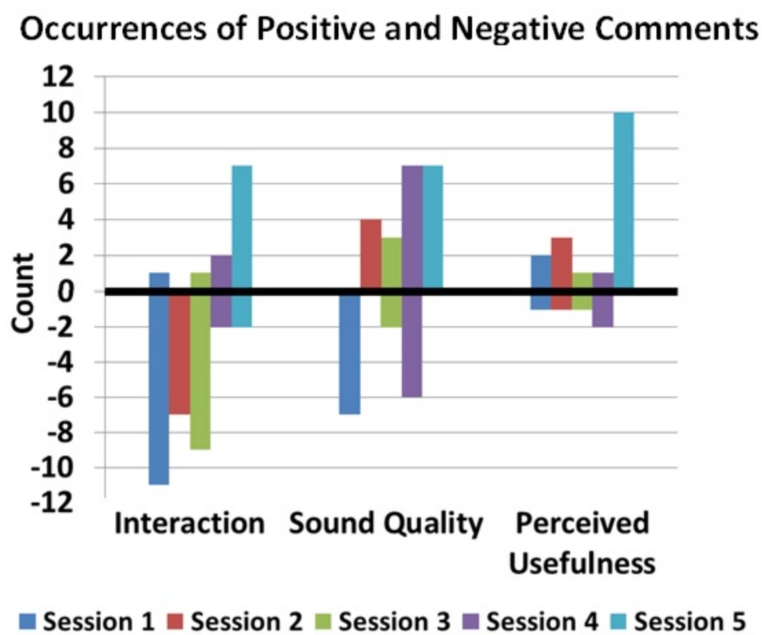

Fig. 7.8: Occurrences of positive and negative comments made under each major category in post-performance discussions.

our efforts on understanding and enhancing user interaction and collaboration in a distributed setting. However, it is unrealistic to expect the average performer using a standard Internet connection to have access to such ideal conditions. Therefore, we sought to assess musician performance with our system under the greater delays typically experienced under less optimal network conditions. In order to maintain the experimental controllability afforded by our laboratory setup, and bypass the problematic nature of carrying out experiments spanning multiple physical locations, we decided instead to simulate increasing levels of latency with our existing setup. As such, during the final session we conducted with the musicians, we repeatedly doubled our baseline latency of 16 ms until it reached 256 ms, placing our system in the realm of the Latency Accepting Approach. Under each level of latency, we asked the musicians to perform two songs, or for approximately ten minutes, while selecting from our list of system features as they saw fit, in a manner similar to the freestyle session described above. Our goal was to determine how well the features fared under increasing latencies, and whether any of them particularly helped the musicians cope with such changes.

Figures 7.9 to 7.13 illustrate the usage patterns for our system features by each of the musicians, under increasing levels of latency. It is worth noting that the absence of track panning and mix control usage at a latency of 32 ms, as seen in Figure 7.10, could be an anomaly, with the musicians perhaps forgetting to turn the features on, or feeling that such features might not be necessary for the specific songs played during that segment. Figure 7.14 depicts the total duration for which each feature was used, as averaged across the three musicians, also under increasing levels of latency. While no particular trends emerged as the latencies increased, dynamic volume and dynamic reverb continued to be the most popular features, on the basis of total duration of use, followed by musician spatialization. As with the results of the freestyle session described above, mix control and track panning were used only intermittently, and less frequently when compared to the other three features.

Our video footage indicates that the musicians were able to perform reasonably well under the 32 ms latency, and while they found the 64 ms latency to be slightly perceptible, they were nonetheless able to play their songs with relative ease. How-

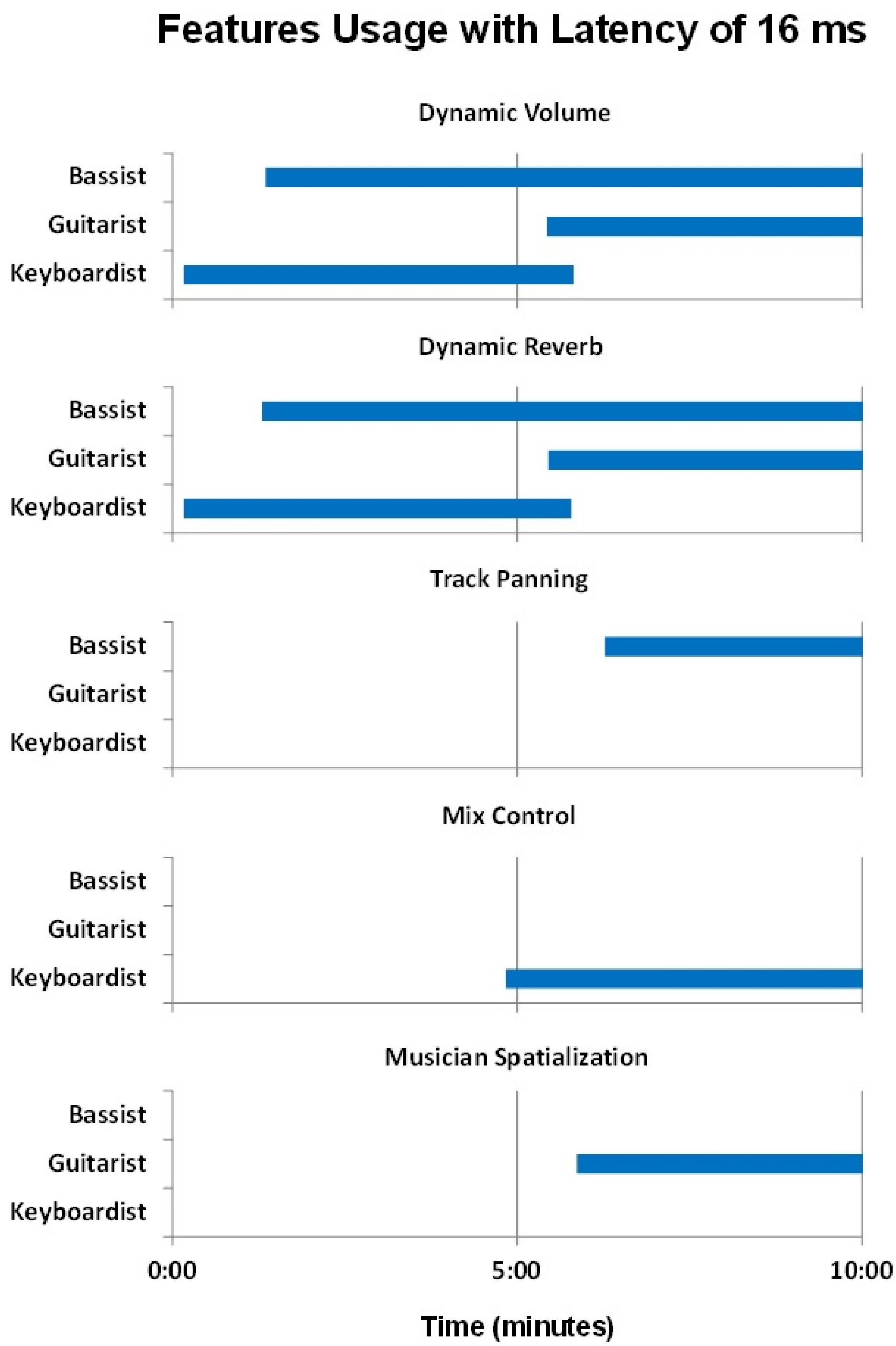

Fig. 7.9: Usage pattern of features by musicians under base latency of 16 ms.

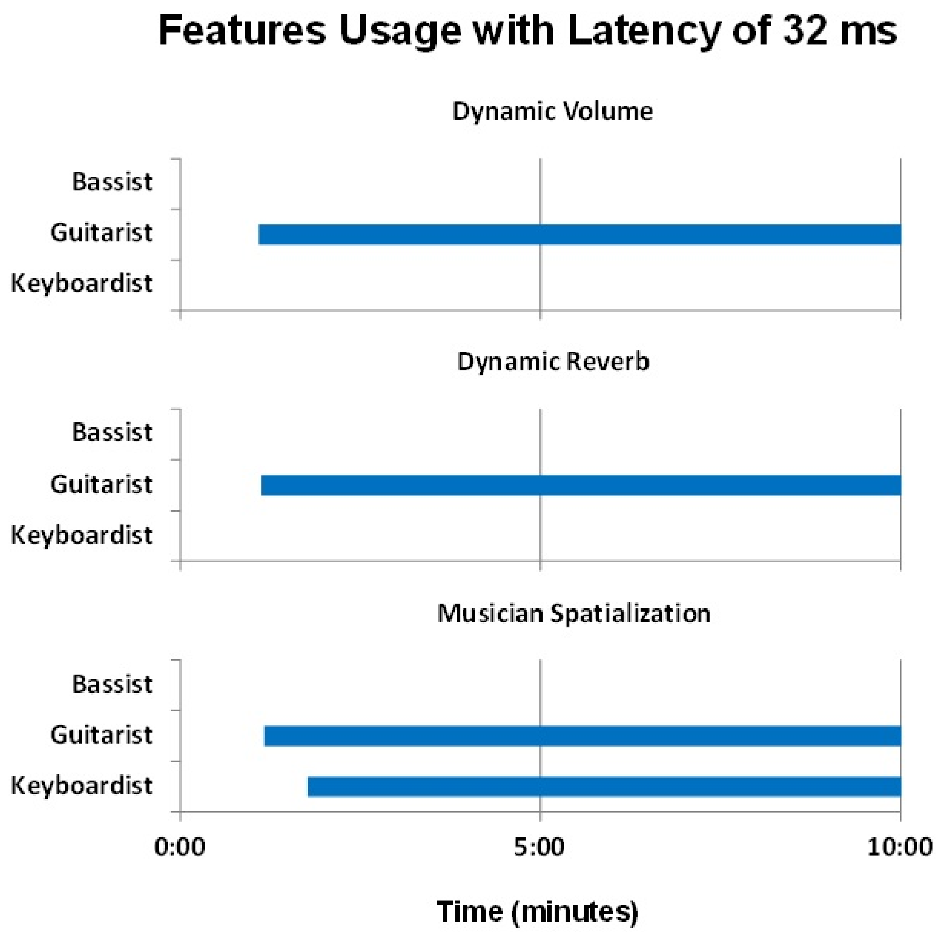

Fig. 7.10: Usage pattern of features by musicians under latency of 32 ms.

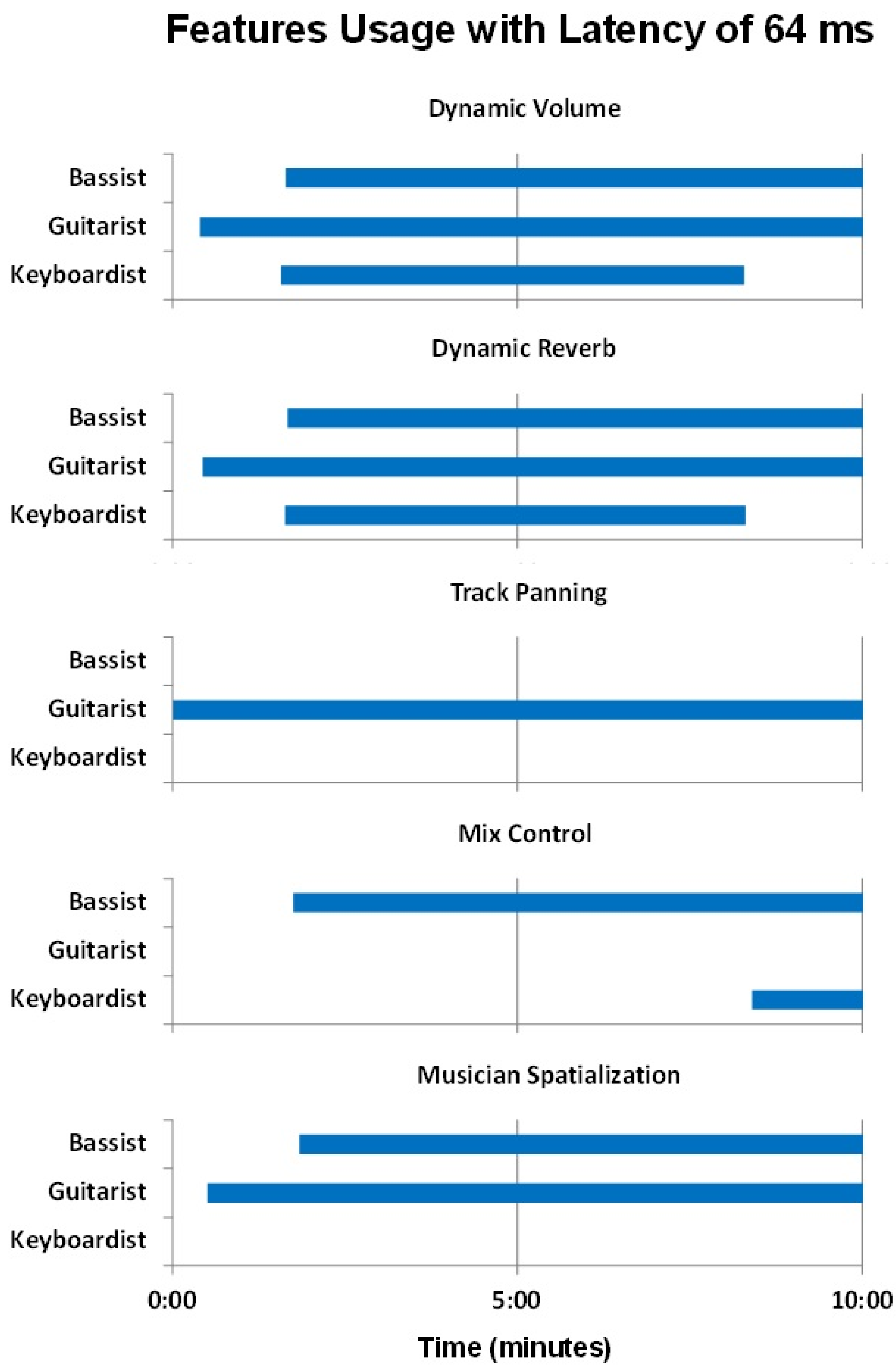

Fig. 7.11: Usage pattern of features by musicians under latency of 64 ms.

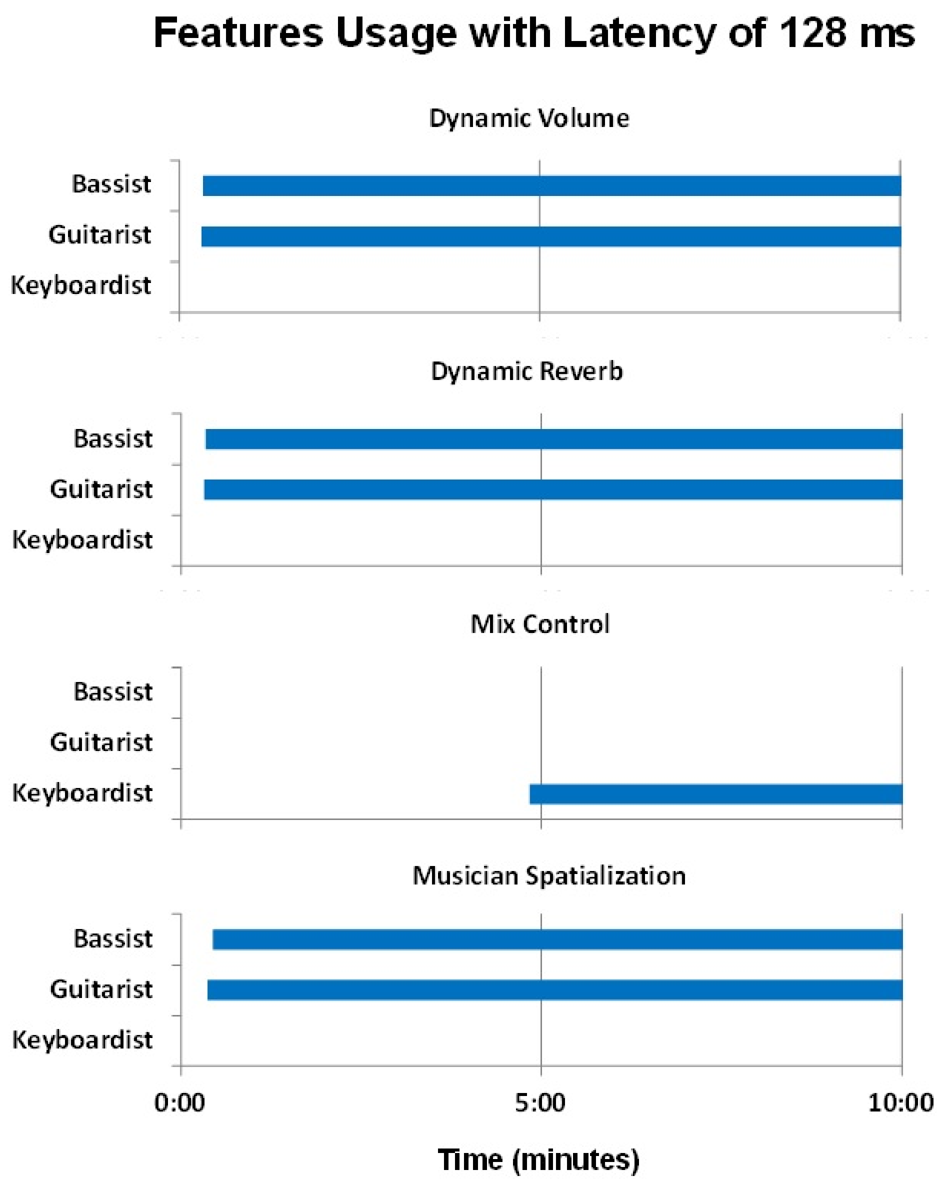

Fig. 7.12: Usage pattern of features by musicians under latency of 128 ms.

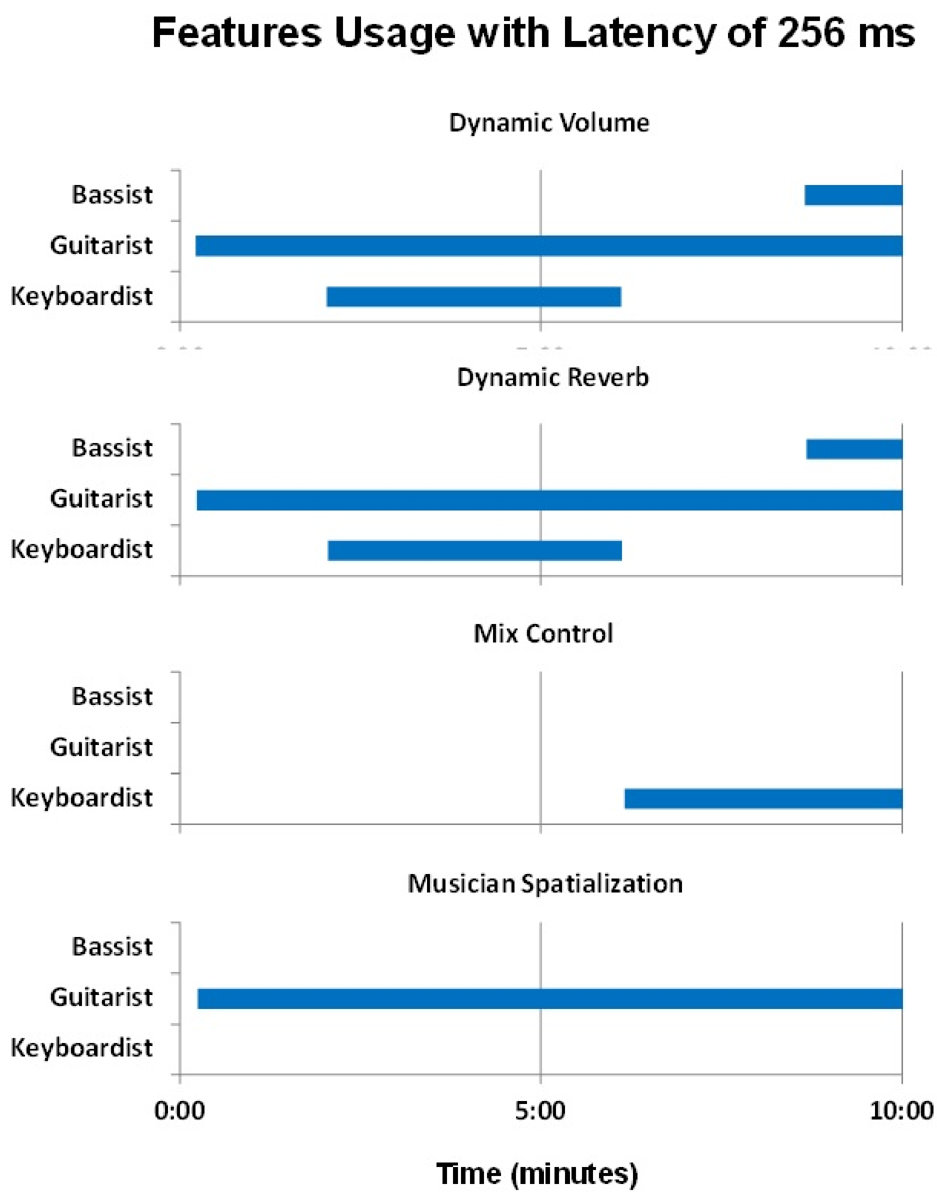

Fig. 7.13: Usage pattern of features by musicians under latency of 256 ms.

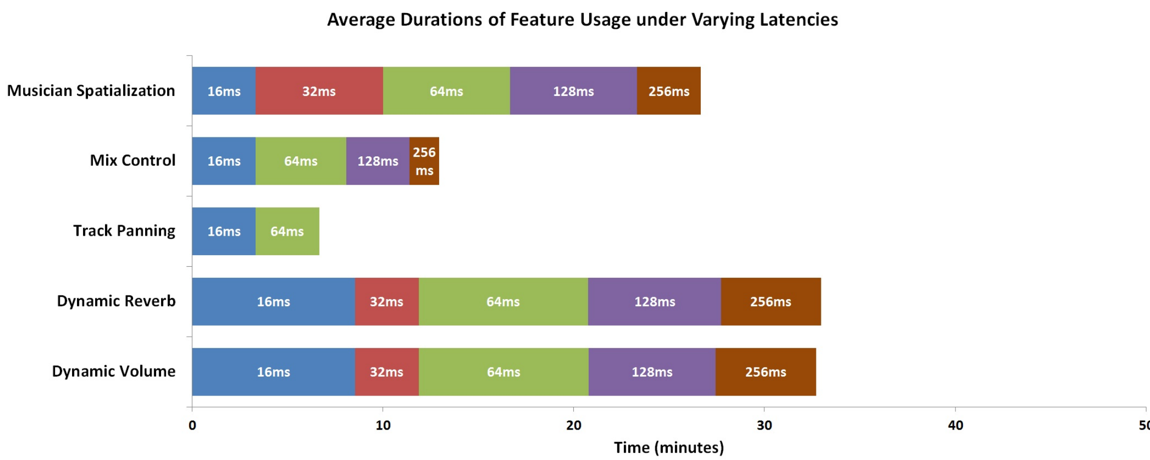

Fig. 7.14: Total length of usage for each feature under various latencies, averaged across the three musicians.

ever, once we doubled the latency to 128 ms and beyond, they found it too difficult to cope with the delays, often breaking mid-song to complain to one another, as they described the effect as "disorienting", "challenging" and even "nauseating". During the post-test discussion, the guitarist revealed the challenge of remaining in time under the greater latencies. While the musicians attempted to "fall back into routine", and simply play the songs as they typically rehearsed them, they were still unable to get through a full song. Such a coping mechanism was also observed by Chew et al., and described earlier in Section 2.2.2 [47]. Thus, with their focus placed entirely on keeping a rhythm, the musicians found it difficult to pay attention to the system features, and instead resorted to simply turning them on or off at various intervals, hoping that they could help alleviate some of their frustrations. This, in turn, explains the lack of trend in the usage patterns described earlier. Nonetheless, all three musicians pointed to the usefulness of the musician spatialization feature, explaining how it facilitated the challenge of keeping time by allowing them to easily distinguish one another's instruments and focus on a rhythm instrument such as the bass whenever necessary, without the need for explicit gestures that might detract from their focus. The keyboardist, for instance, explained:

"The spatialization one I liked. I could separate [the guitarist's name] from [the bassist's name], and then try to keep up time with [the bassist's name]'s bass."

In addition, dynamic volume and dynamic reverb continued to be among the guitarist's and bassist's favourites, with the former describing how, in contrast, the gestures necessary to operate the track panning and mix control features made them "distracting". This corresponds with our observations during the freestyle session, described in Section 7.2.5 above, which indicated that such features might require a greater cognitive load than the other three, an issue likely aggravated when coupled with the challenge of coping with added latency.

Chapter 8

Participatory Design Cycle

Our long-term deployment with the three-piece band was beneficial in allowing us to fine-tune our system and introduce new features. Nonetheless, while we held several sessions with the band, each focused on a different aspect of the system. As such, the musicians had a relatively limited exposure to our responsive environment for distributed performance as a whole. Furthermore, although we strove to maintain ecological validity by creating a relaxed environment where musicians were asked to play several songs of their choice without interruptions, experimental settings can only be considered naturalistic to a certain extent. Finally, while the A/B/A tests allowed us to isolate the effects of each feature on distributed performance, it was not necessarily indicative of how musicians might experiment with our system on their own terms.

As a result, we organized an "artist residency" with Steve Cowan,[1] an award-winning performer, composer and teacher. Mr. Cowan has had a passion for music from a very young age, having taken piano lessons from the ages of 6 to 10, followed by drum lessons at the age of 11. In high school, the guitar became his main focus, and remains as such to this day. After earning his Master of Music degree from the Manhattan School of Music, Mr. Cowan moved to Montreal, where he continues to be an active member of the city's music scene, teaching guitar and performing leisurely

[1]http://about.me/stevecowan

with a number of different bands in his spare time. Therefore, he fit our “Rocker” persona quite well, and was a suitable candidate for this participatory design cycle.

Our goal was to examine how well the features we had designed, both through our early efforts and the long-term deployment, lent themselves purely to the process of making music, regardless of the context in which they were used (i.e., distributed vs. co-present). In addition, we hoped to examine their potential for creative engagement, a quality that is difficult to measure accurately in an experimental setting, and one that did not particularly improve with the use of our features (dynamic volume being the exception) in the A/B/A tests described earlier. Finally, we were particularly interested in examining the difference in depth, quality and nature of the feedback that would emerge from a participatory design process, in comparison with the results iteratively drawn from experiments under the standard user-centered design process, and employed throughout our developments thus far.

Mr. Cowan was asked to write a few musical pieces using our responsive environment. However, we specifically clarified that the final compositions themselves were not necessarily the most critical outcome to our research. Rather, they served as a vehicle to help us understand the extent to which our system features could support and, through his recommendations, perhaps even improve the creative process. As such, Mr. Cowan was given a very active role, one on equal footing with the system designers, and explicitly informed that his criticisms and suggestions, no matter how extensive, would play a crucial part in shaping any further iterations of the responsive environment for distributed performance.

8.1 Methodology

Given that our prototype had already undergone multiple iterations, the participatory design technique most suitable for our needs was that of cooperative prototyping, which entails delivering a system to its end-users as a series of iterative prototypes, each of which gradually adds functionality [138]. As Muller explains, cooperative prototyping offers several advantages, including enhanced communication by grounding discussions in concrete artefacts, and improved working relations through a sense of shared ownership of the resulting system. Furthermore, unlike other participatory design techniques typically held at the early stages of design, such as workshops or storyboarding, cooperative prototyping allows designers to benefit from an understanding of the constraints posed by the practical limitations of the software. In turn, this makes it particularly suited to the design of creative or artistic systems, where the user tasks, needs or goals needed to guide and ground discussions may prove difficult to define during the initial stages of a project.

According to Muller, the success of this technique hinges on presenting each prototype as a "crucial artifact in the end user's work" [138], which allows them to form ecologically valid impressions of the system. In turn, this creates the basis for an on-going conversation between designers and users about the changes that could potentially improve work practices and their technical feasibility, all while finding a balance as both parties challenge one another's assumptions.

As such, our collaboration with Mr. Cowan lasted 14 weeks, with sessions being held on a regular basis every 1-2 weeks. Seeing as the composer would be using the system on his own, it was necessary to create an environment where we could simulate interactions with virtual remote participants. Mr. Cowan therefore used a Loop Station, a floor-based device that allowed him to record and play back multiple tracks on the fly. Output channels from the Loop Station were routed to input channels on our Edirol FA-101 capture interface and, in turn, treated by our SuperCollider software as though they were separate audio streams created by different musicians, and subsequently layered into one mix. In addition, the audio output channels from our SuperCollider software were connected to input tracks in Ardour, allowing Mr. Cowan effectively to record his compositions to a laptop. Audio routing between the capture interface, SuperCollider and Ardour was implemented via JACK, and a Microsoft Kinect served to track Mr. Cowan's position and orientation. The modified system configuration can be seen in Figure 8.1.

In addition, Mr. Cowan participated in the design of a modified interface that would offer him complete control over the virtual remote participants from one station. As such, Mr. Cowan was able to use the updated main GUI (seen in Figure 8.2) to set the base volume and reverb levels for himself and other virtual participants at

Fig. 8.1: System configuration used throughout participatory design cycle.

the very start of a session, and the new secondary GUI (seen in Figure 8.3) to "move" the virtual musicians' avatars and select features. Moving the avatars allowed him to experiment with the subset of the overall dynamic volume and dynamic reverb ranges he stood to experience. Specifically, the range for both features is determined as a function of the minimum and maximum possible distances between any two avatars. Thus, for example, moving one band member's avatar significantly closer to his own reduces the maximum distance that can be achieved relative to that avatar as he moves about his physical space. As a result, the composer could experience a subset of volume changes closer to the higher end of the possible dynamic volume range, and a subset of reverb changes closer to the lower end of the possible dynamic reverb range for that particular avatar.

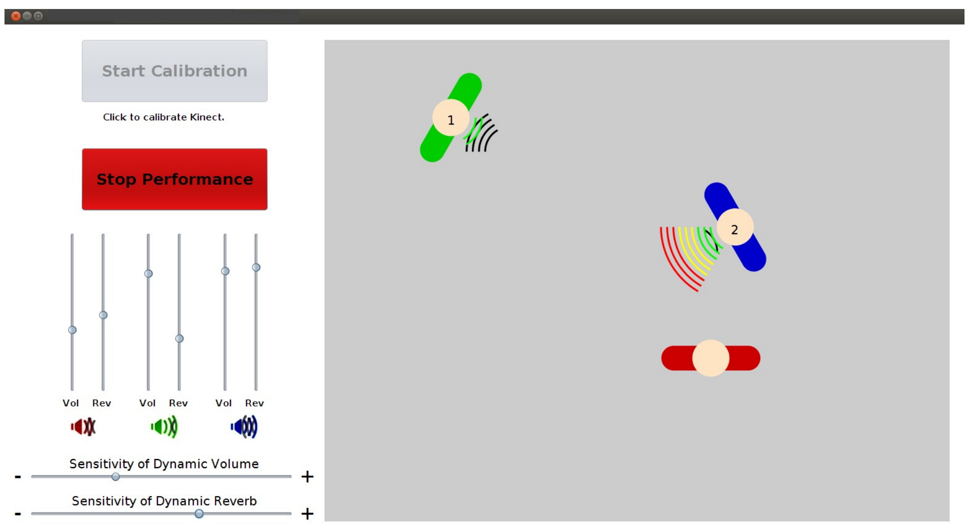

Fig. 8.2: Composer's main GUI.

Mr. Cowan spent the first few sessions familiarizing himself with the system, and determining how to best approach the task that was assigned to him. After this introductory phase, he began shifting his focus towards experimentation . As such, each session would begin with a discussion of any changes made to the system as a result of previous suggestions. Subsequently, Mr. Cowan would spend a few hours playing music and interacting with the system. During this exploratory portion of the

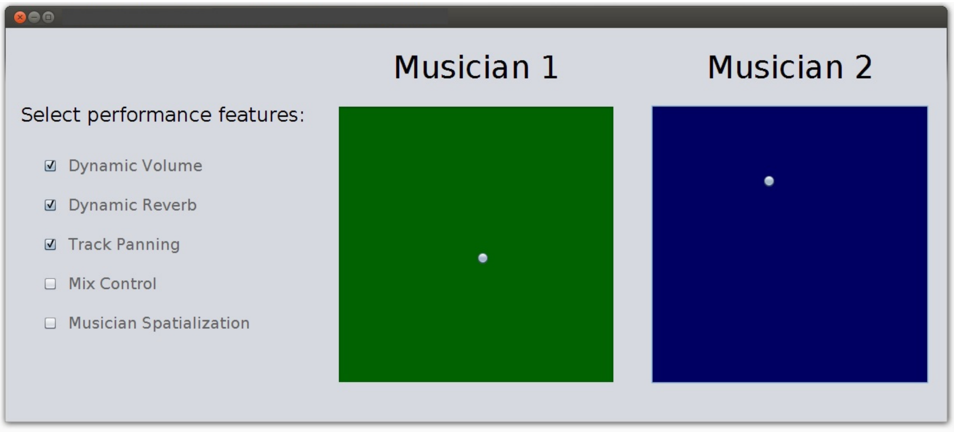

Fig. 8.3: Composer's secondary GUI.

session, Mr. Cowan would typically record his impressions in point-form notes, while we provided our assistance on demand, and only in a technical capacity, for instance, helping resolve any glitches with the system or providing clarifications when needed. Afterwards, a discussion would be held, allowing Mr. Cowan to share the notes he had made, and describe how the system could be improved for the following week's session. The composer would then take a few days to expand on the ideas contained in his notes, before sending us a full report that typically included additional details and explanations for his recommendations, and comments on the progress of the pieces thus far. In the final weeks, as Mr. Cowan determined the system to have reached a satisfactory state, and with fewer recommendations to make, he began to immerse himself fully in the process of composition.

8.2 Outcomes

Seeing as our system had already undergone multiple design iterations, we did question whether we would be able to reap the full benefits of participatory design, a process typically held during the early stages of development. As such, our expectations going into this collaboration were that Mr. Cowan would help us further "tweak" the system, and comment on how useful the features were to the process of composition. However, somewhat surprisingly, the outcome of the collaboration exceeded our expectations on both counts. First, in addition, to making recommendations for improving existing system features, Mr. Cowan was the source behind new additions to the system. For instance, he introduced the idea behind the mix control feature, as he believed it could capitalize on head tilting motions better than track panning, which, as described earlier, proved an unpopular mapping with the musicians during the long-term deployment tests. Mr. Cowan reasoned that head tilting would be better suited to attentive listening, as musicians often do in a studio setting, as they lean their heads into one headphone at a time. In addition, he helped us shape the reverb feature by designing the effect itself, and selecting its minimum and maximum value. He also made recommendations to the design of the graphical avatars, suggesting that "echo" waves be used to denote reverb levels, and that sound waves denoting volume should indicate when the maximum level is reached (both modifications can be seen in Figure 8.2). What is perhaps even more interesting, however, is that Mr. Cowan began combining the features in ways we had not anticipated. For instance, he would often isolate one of the tracks using the track panning feature, then use the dynamic volume and dynamic reverb to experience variations in volume and reverb exclusively on that track. He also found that the usefulness of each feature and their possible combinations varied throughout the different stages of the compositional process, a topic he discusses in further detail in the following section.

8.2.1 The Composer's Report

After the final session, we requested that Mr. Cowan write a full report summarizing his experience. In particular, we asked that he reflect on the process of collaboration itself, his impressions of the final system, its effects on musical composition, and whether he would consider using it in the future. Overall, Mr. Cowan found that embodied interactions lent themselves particularly well to seamless experimentation with various mix settings, which, in turn, helped facilitate the process of composition. He explained that he previously had a tendency to avoid the post-composition mixing process:

“Almost every musician I know these days has some sort of recording software on their computer, and thus has the ability to record and produce multi-track recordings at home. Personally, I find all the clicking and computer-based activity in this to drain my creative energy and make the process frustrating.”

In contrast, however, he found the ability to compose and mix simultaneously to be particularly beneficial:

“Using the performance system here, I was able to get some great solutions for these issues without having to do anything other than play my music in real time, and move my body a bit. I was easily able to see which tracks sounded best panned left, or right, or in the center; I was able to hear which textures were better off in the foreground, and which sounded better off more “distant”, perhaps with a hint of reverb; I was able to iron out how two musical ideas interacted one on one, and then with a slight 90 degree turn, could hear how it then sounded with a third musical idea in the mix.”

Mr. Cowan further detailed how certain features proved to be particularly well-matched to specific stages of the compositional process:

“Other than dynamic manipulations to volume and reverb, the three features I worked with also provided a logical succession for the creative process. Track panning allows the ability to work on ideas one on one, by cutting out one of the 3 musicians with a simple torso pivot. The mix control brings all 3 players into the mix, but with the ability to pan your own part around to see how everything is blending/working together. Then the spatialization is a good final step, fleshing out the music ideas into their own space within the panning, and hearing how it works in a situation that will sound closer to the eventual desired final product (be it a live performance or a recording).”

In summary, he had a positive impression of the overall system:

“In conclusion, the features that this system offered were fun, useful, and helped me come up with new musical and production ideas.”

However, he also offered important criticisms, explaining, for instance, that the system’s current motion tracking technique may prove inadequate for instruments that require musicians to be seated, such as the keyboard. Furthermore, he anticipated that the lack of precise, numerical representation of the various levels effected by the system features might make it more difficult to correctly re-create the mix when working on the final, polished product.

The entire report, in its original form, can be found in Appendix A.

8.2.2 Reflections on the Participatory Design Cycle

In the end, we found that holding a participatory design cycle this late in the development process to be rather challenging. Given the nature of its position in the development chain, cooperative prototyping requires designers to become receptive towards changing a systems to which they have, perhaps inevitably, grown quite attached. Thus, having already invested considerable efforts in prototyping our system, receiving and accepting extensive criticism was a daunting process. A key to the success of our collaboration with Mr. Cowan was in regarding his role as that of a true expert, one whose decisions and influence should be placed on equal footing with those of the system designers. In turn, it was by allowing Mr. Cowan to take on such an active role that we discovered that the usefulness of our system features extended beyond their intended context of distributed performance, and into the realm of mixing and composition. Thus, by combining the system modifications designed for Mr. Cowan with our improved features, the participatory design cycle resulted in a new and unexpected artefact: a responsive environment for musical composition. The following section highlights how this new artefact compares with existing tools for mixing and recording.

8.2.3 Comparisons with Existing Solutions

Musical performance and mixing have traditionally been treated as separate processes, which is natural since musicians can hardly be expected to step over repeatedly to a mixing console or computer in order to adjust settings mid-performance. The exception, perhaps, is the case where the computer is also the instrument. We use the term “mixing” to denote “the adjustment of relative volumes, panning and other parameters corresponding to different sound sources, in order to create a technically and aesthetically adequate sound sum" [41]. Digital audio workstations (DAWs) continue to be the gold standard for audio recording, editing and mixing, with possibilities that range from simple two-channel editors to complete recording suites, and include both hardware and software components. However, the vast majority of stations continue to operate according to the same "multitrack tape recorder" metaphor, utilizing mixing consoles that allow musicians to control multiple channels—each carrying an audio track—through pan pots, faders and sliders, or software solutions that simply simulate such mixing consoles.

The drawbacks to such traditional mixing technology are that it significantly constrains composition activities that wish to mix musical input as it is being generated, and its requirement of hands-on interaction is ill-suited to supporting musicians who wish to exercise independent control over their mix during performance. In fact, in spite of the tremendous potential afforded by the advent of digital audio, mixing interfaces have changed very little in the decades following their introduction [41, 125]. As exemplified through such systems as Avid Technology's Pro Tools, Apple's Logic Pro, Ableton Live and Steinberg's CueBase, the software systems most commonly used by professionals and amateurs alike take their inspiration from the mixing console: faders, knobs and sliders are considered standard tools for mix control [73]. However, although a number of systems have sought to facilitate or improve the mixing process through novel solutions, most continue to reflect the console analogy. For instance, while the Lemur2 and Dexter interfaces, both developed by JazzMutant, offer multi-touch to allow users to take advantage of common pinching and expansion gestures for added precision, their layout still emulates that of the mixing console [41, 160]. As another example, the Cuebert system, which also utilizes a multi-touch interface to allow for flexible display of dynamic and context-sensitive content in the "high-pressure" environment of musical theatre, relies on a traditional mixing board paradigm as well [125].

Nonetheless, a few alternatives have been proposed. For instance, Pachet et al. introduced the concept of "dynamic audio mixing", which offers listeners direct control over the spatialization of musical pieces [146]. To facilitate this process, while allowing users to move more than one sound source at a time, the authors employ a constraint paradigm that aims to preserve the properties of the configuration of sound sources that need to be satisfied in order to maintain "coherent, nice-sounding mixings". Such ideas were implemented through MusicSpace, a system whereby speaker icons representing sound sources, and an avatar representing the listener, can be moved graphically to induce real-time changes in the spatial arrangement of an overall piece [147]. This work can also be seen as an example of the emerging active music listening paradigm, which gives listeners the ability to mix and manipulate the different constituent sources, or "stems", of a musical piece on their own [175]. Similarly, Carrascal et al. developed an interface that allows its users to manipulate spatially arranged sound sources, in an attempt to take into account modern mixing technologies such as surround and 3D audio [41]. As another instance of alternative mixing techniques, the waveTable is a tabletop audio waveform editor that combines multi-touch and tangible interaction techniques, allowing users to manipulate sound samples directly [162]. Furthermore, the Chopping Board allows users to "chop" and re-sequence tracks through interaction with a physical "editing pad" that can detect their gestures through a combination of infrared and touch sensors [122]. Another example is Noisescape, a 3D first-person computer game where users can collaboratively compose complex musical structures, by creating and combining elements with varying physical attributes [83]. Finally, as previously detailed in Section 2.3.4, the Sound Maker was designed to map a user's location and movement to changes in the pitch, tempo and volume of an electronically-generated percussive stream. However, much like those inspired by mixing consoles, the systems described here do not support simultaneous performance with an instrument and mixing by the same user.

In contrast, the artefact that resulted from our collaboration with Mr. Cowan allows for hands-free, seamless, dynamic control of musical parameters during performance through the use of embodied interaction. Our responsive environment for musical composition allows musicians to mix a live instrument simply by moving around their space. As Mr. Cowan noted, the ability to play and mix instruments simultaneously helped enhance the creative process, allowing him to experiment with various settings in real-time and, in turn, determine how an instrument might best be integrated into a final piece. We note that our artefact is not necessarily meant for producing polished, final works. Rather, it can help musicians experiment seamlessly with various mix possibilities during the process of composition, in order to determine how an instrument might best fit among others in the final recording. Nonetheless, we believe that our responsive environment for musical composition could prove beneficial for musicians seeking alternatives to traditional mixing solutions that may enhance their creativity during composition.

Chapter 9

Release Candidate

The release candidate for the responsive environment for distributed performance represents the culmination of all the design and development efforts described in this dissertation thus far. Seen in use by a musician in Figure 9.1, the system evolved from a simple idea to augment collaboration within the context of distributed music, to an environment that offers musicians seamless, hands-free control over their instrumental mix by capitalizing on simple, common interactions. Table 9.1 describes the importance of each development phase in helping shape the release candidate. Furthermore, Figure 9.2 depicts the entire user-driven evolution of our system, illustrating the origin of the various system features, and the impact of each stage of user involvement.

9.1 System Features

The release candidate encompasses five features in total: dynamic volume, dynamic reverb, mix control, track panning and musician spatialization, each of which we detail and formalize in this section.

Phase	Outcomes
User Observations	• List of user behaviours • Personas
User interviews	• List of performance criteria
Early Prototypes	• Introduction of **dynamic volume** in a co-present setting • Formalization of performance criteria and corresponding evaluation techniques
Alpha System	• Distribution of system across three locations • Introduction of **track panning** feature • Implementation of graphical user interface • Design of dynamic graphical animations
Beta System	• Introduction of **dynamic reverb** feature • Introduction of **musician spatialization** feature
Participatory Design	• Introduction of **mix control** feature • Development of a responsive environment for musical composition

Table 9.1: Summary of the various project phases and the major project features introduced by each.

Fig. 9.1: A musician interacting with the responsive environment for distributed performance.

9.1.1 Dynamic Volume

The dynamic volume feature allows musicians to affect each other's volume levels as follows: as one musician moves *towards* another's virtual location, both can experience each other's instruments as gradually increasing in volume. The converse holds true as a musician moves *away* from another's virtual location. To formalize this feature, assume that M musicians located at virtual locations $\vec{p_i} \in \mathbb{R}, i = 1, \ldots, M$ are each producing the source audio signal $s_i(t)$. Each musician i sets their initial base volume, v_i^{base}. As such, all musicians receive an overall base mix of

$$m^{base}(t) = \sum_{j=1}^{M} v_j^{base} s_j(t) \tag{9.1}$$

We calculate d_{ij}^{max}, a maximum possible distance between musicians i and j based on the sizes of their local spaces. Note that the size of the local spaces can be determined during the Kinect calibration phase, which each musician can carry out independently, as will be explained later in this chapter. Subsequently, we define a minimum threshold distance, t_{ij}^{min}, and a maximum threshold distance, t_{ij}^{max}, selected as proportions of d_{ij}^{max}. When dynamic volume is in use, the new mix heard by each

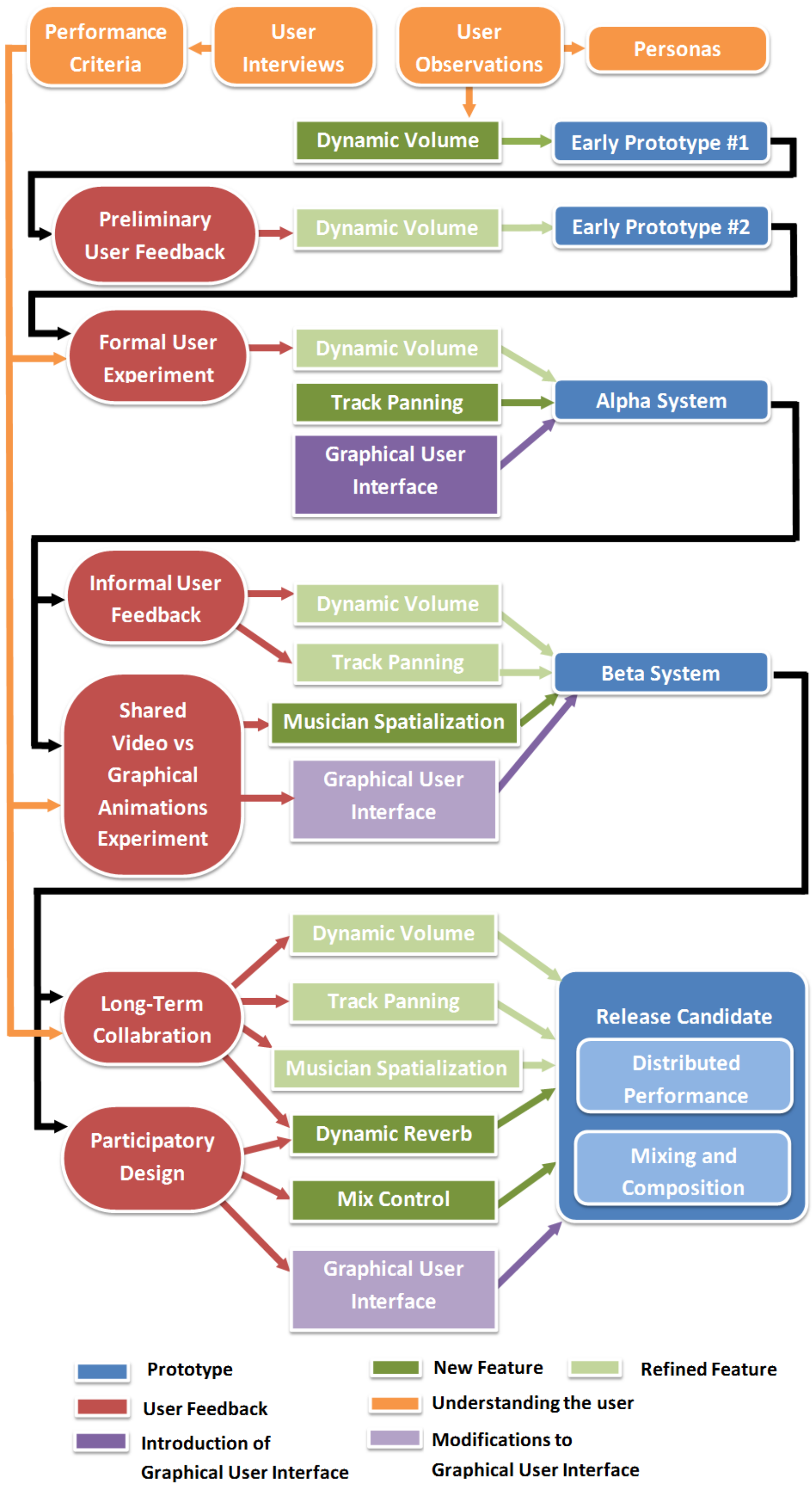

Fig. 9.2: Complete user-driven evolution of the responsive environment for distributed performance.

musician i is calculated as

$$m_i(t) = v_i^{base} s_i(t) + \sum_{j \neq i}^{M} v_{ij} s_j(t) \tag{9.2}$$

where the volume of musician j as perceived by musician i is defined as

$$v_{ij} = f(\|\vec{p}_i - \vec{p}_j\|) \tag{9.3}$$

and f is a monotonically decreasing exponential function that maps $t_{ij}^{min} \leq \|\vec{p}_i - \vec{p}_j\| \leq t_{ij}^{max}$ to $v_{ij}^{min} \leq v_{ij} \leq v_{ij}^{base}$, where v_{ij}^{min} is a value that is inversely proportional to the dynamic volume sensitivity chosen by each musician i. We note that since f is an exponential function, we can model a linear increase on the decibel scale that matches the musicians' expectations. A graphical representation of the feature can be seen in Figure 9.3

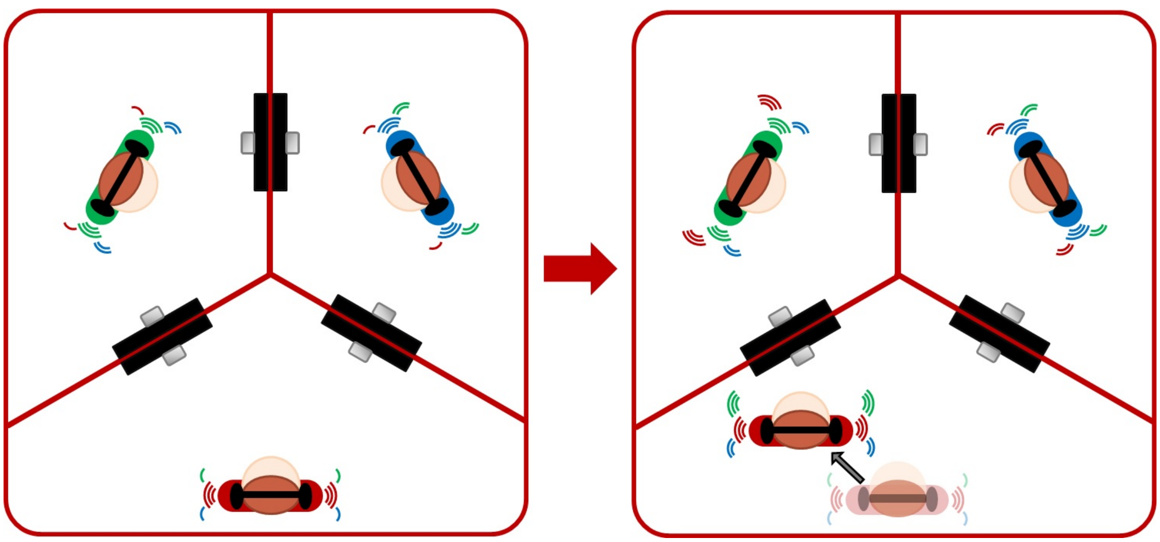

Fig. 9.3: Graphical respresentation of the dynamic volume feature.

9.1.2 Dynamic Reverb

The dynamic reverb feature allows musicians to affect each other's reverb levels as follows: as one musician moves *away* from another's virtual location, both can experience each other's instrument sounds as gradually increasing in reverberation. The converse holds true when a musician moves *towards* another's virtual location. To formalize this feature, assume that M musicians located at virtual locations $\vec{p_i} \in \mathbb{R}, i = 1, \ldots, M$ are producing each the source audio signals $s_i(t)$. Each musician i sets their initial base reverb r_i^{base} , and their volume v_i. As such, all musicians receive an overall base mix of

$$m^{base}(t) = \sum_{j=1}^{M}(v_j s_j(t) + r_j^{base}(t)) \tag{9.4}$$

Reverb for each musician, r_i, is calculated as

$$r_i(t) = \texttt{Gverb}((s_i(t), rt_i, d_i)) \tag{9.5}$$

where `GVerb` is a built-in SuperCollider reverb function that requires the reverberation time in seconds, rt, and a damping value between 0 and 1, d, as its inputs. We calculate d_{ij}^{max}, a maximum possible distance between musicians i and j based on the sizes of their local spaces. Subsequently, we define a threshold distance, t_{ij}, selected as a proportion of d_{ij}^{max}. When dynamic reverb is in use, the new mix heard by each musician i is calculated as

$$m_i(t) = v_i s_i(t) + r_i^{base}(t) + \sum_{j \neq i}^{M}(v_j s_j(t) + r_{ij}(t)) \tag{9.6}$$

where the reverb of musician j as perceived by musician i is defined as

$$r_{ij} = g(\|\vec{p_i} - \vec{p_j}\|) \tag{9.7}$$

and g is a monotonically increasing linear function that maps $t_{ij} \leq \|\vec{p}_i - \vec{p}_j\| \leq d_{ij}^{max}$ to $0 \leq d_{ij} \leq 1$ and $0 \leq rt_{ij} \leq rt_{ij}^{max}$. The value of rt_{ij}^{max} is directly proportional to the dynamic reverb sensitivity chosen by each musician i. A graphical representation of the feature can be seen in Figure 9.4

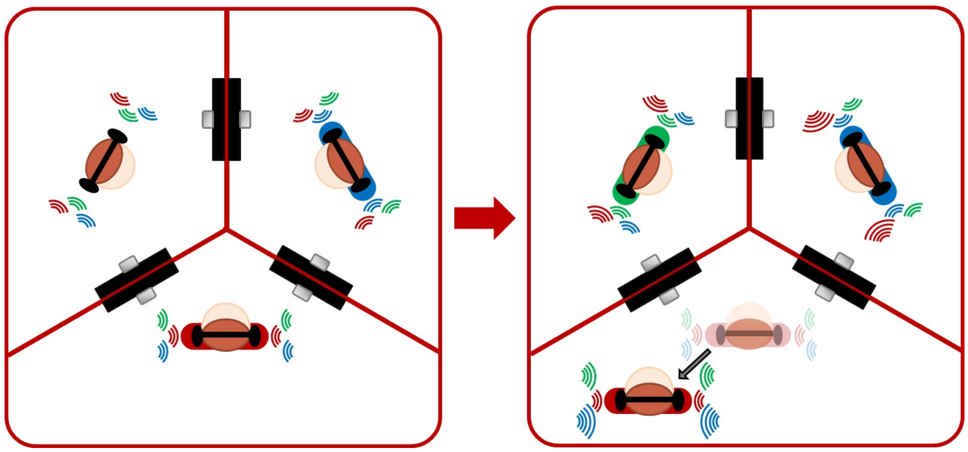

Fig. 9.4: Graphical representation of the dynamic reverb feature.

9.1.3 Mix Control

The mix control feature allows each local musician to change the mix of his instrument with those of the remote musicians by tilting his head. Tilting to the left will move the sound of his instrument, along with that of the musician whose virtual location is to his left, entirely through the left headphone. The instrument of the musician whose virtual location is to his right will be heard unaccompanied through the right headphone. The converse holds true when the local musician turns his head to the right. To formalize this feature, assume that M musicians located at virtual locations $\vec{p}_i = [x_i, y_i] \in \mathbb{R}, i = 1, \ldots, M$ are each producing the source audio signals $s_i(t)$. Each musician i sets a base volume of v_i^{base} and receives an overall mix $\vec{m}_i = (m_{Li}, m_{Ri})$, representing the left and right audio channels. As such, all musician receives an overall mix

$$m_k(t) = \sum_{j=1}^{M} \frac{v_j^{base}}{2} s_j(t) \tag{9.8}$$

where $k \in \{L\text{=left}, R\text{=right}\}$. When mix control is in use, each musician i receives a mix

$$m_{ki}(t) = \sum_{j=1}^{M} a_{kij} \frac{v_j^{base}}{2} s_j(t) \tag{9.9}$$

where a_{kij}, a channel-wise mixing coefficient that depends both on the relative positions of musicians i and j, and the head roll of musician i, $|\phi_i| \leq \phi_{max}$, is defined as follows:

- Let $b_L = -1, \quad b_R = 1$.
- If $i = j$, or if $sgn(\phi_i) = sgn(x_i - x_j)$ (i.e. musician i tilts his head towards musician j) , then

  $$a_{kij} = \begin{cases} h(\phi_i), & \text{if } sgn(\phi_i) = sgn(b_k) \\ f(\phi_i), & \text{if } sgn(\phi_i) \neq sgn(b_k) \end{cases} \tag{9.10}$$

  where h is a monotonically increasing exponential function that maps $\phi \in b_k[0, \phi_{max}]$ to the range [1,1.25] (the upper limit of the range being a value selected by Mr. Cowan), and f is a monotonically decreasing exponential function that maps $\phi \in b_k[0, \phi_{max}]$ to the range [0,1]. Such a mapping allows the remote musician towards whom the local musician tilts his head to come across as slightly louder through the headphone corresponding to the direction of tilt, and significantly quieter through the opposite headphone.
- On the other hand, if $sgn(\phi_i) \neq sgn(x_i - x_j)$ (i.e. musician i tilts his head

away from musician j), then

$$a_{kij} = \begin{cases} f(\phi_i), & \text{if } sgn(\phi_i) = sgn(b_k) \\ 1, & \text{if } sgn(\phi_i) \neq sgn(b_k) \end{cases} \tag{9.11}$$

where f is the same exponential function described above. Such a mapping allows the remote musician away from whom the local musician tilts his head to come across as significantly quieter through the headphone corresponding to the direction of tilt, while remaining the same through the opposite headphone.

We note that the value of ϕ_{max} is inversely proportional to the sensitivity of the mix control feature: a greater value requires the musicians to tilt their heads further in order to experience the full effect. Through trials with the composer and the musicians who participated in the long-term deployment, we determined that setting ϕ_{max} at 22.5 degrees proved satisfactory. A graphical representation of the mix control feature can be seen in Figure 9.5.

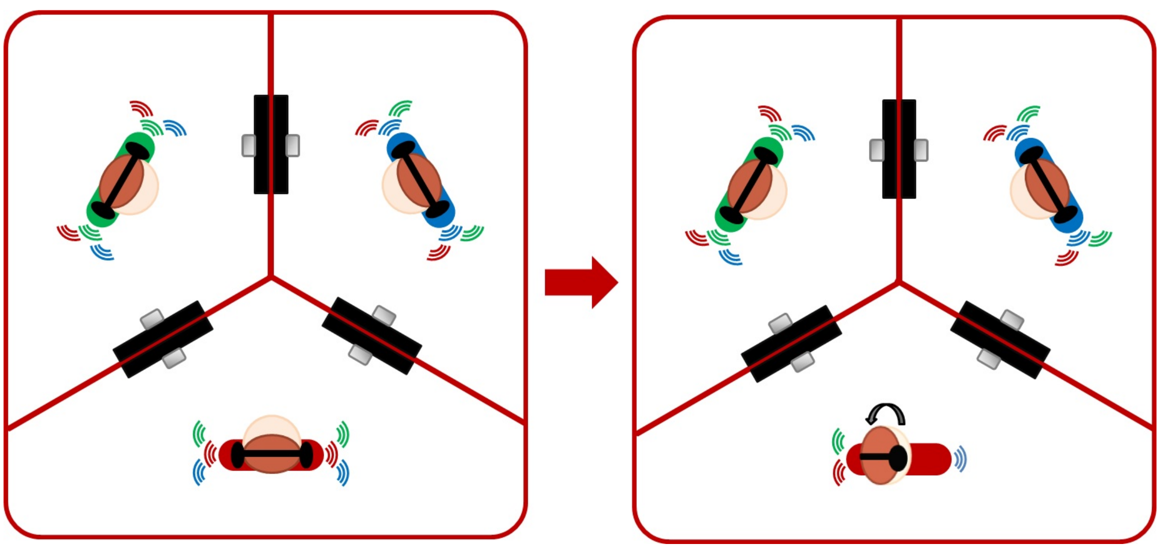

Fig. 9.5: Graphical representation of the mix control feature.

9.1.4 Track Panning

A local musician can isolate each of the tracks of the remote musicians by changing his body's orientation. Turning his body to the left will allow him to hear only the instrument of the musician whose virtual location is to his left, entirely through the left headphone. The instrument of the musician whose virtual location is to his right will become silent. The local musician's own instrument will continue to sound the same, coming through both headphones. The converse holds true when the local musician turns his body to the right. To formalize this feature, assume that M musicians located at virtual locations $\vec{p_i} = [x_i, y_i] \in \mathbb{R}, i = 1, \ldots, M$ are each producing the source audio signals $s_i(t)$. Each musician i sets a base volume of v_i^{base} and receives an overall mix $\vec{m}_i = (m_{Li}, m_{Ri})$, representing the left and right audio channels. As such, all musician receives an overall mix

$$m_k(t) = \sum_{j=1}^{M} \frac{v_j^{base}}{2} s_j(t) \tag{9.12}$$

where $k \in \{L\text{=left}, R\text{=right}\}$. When track panning is in use, each musician i receives a mix

$$m_{ki}(t) = \frac{v_i^{base}}{2} s_i(t) + \sum_{j \neq i}^{M} a_{kij} \frac{v_j^{base}}{2} s_j(t) \tag{9.13}$$

where a_{kij}, a channel-wise mixing coefficient that depends both on the relative positions of musicians i and j, and the body yaw of musician i, $|\theta_i| \leq \theta_{max}$, is defined as follows:

- Let $b_L = -1, \quad b_R = 1$.
- If $sgn(\theta_i) = sgn(x_i - x_j)$ (i.e. musician i turns his body towards musician j), then

$$a_{kij} = \begin{cases} h(\theta_i), & \text{if } sgn(\theta_i) = sgn(b_k) \\ f(\theta_i), & \text{if } sgn(\theta_i) \neq sgn(b_k) \end{cases} \tag{9.14}$$

  where h is a monotonically increasing exponential function that maps $\theta \in b_k[0, \theta_{max}]$ to the range [1,1.25] (a value again chosen by Mr. Cowan), and f

is a monotonically decreasing exponential function that maps $\theta \in b_k[0, \theta_{max}]$ to the range [0,1]. Such a mapping allows the volume of the remote musician towards whom the local musician turns his body to come across as slightly louder through the headphone corresponding to the direction of the turn, and significantly quieter through the opposite headphone.

- On the other hand, if $sgn(\theta_i) \neq sgn(x_i - x_j)$ (i.e. musician i turns his body away from musician j), then

$$a_{kij} = f(\theta_i) \tag{9.15}$$

where f is the same exponential function described above. Such a mapping allows the volume of the remote musician away from whom the local musician turns his body to come across as significantly quieter through both headphones.

We note that the value of θ_{max} is inversely proportional to the sensitivity of the track panning feature: a greater value requires the musicians to turn their bodies further in order to experience the full effect. Through trials with the composer and the musicians who participated in the long-term deployment, we determined that setting θ_{max} at 22.5 degrees proved satisfactory. A graphical representation of the track panning feature can be seen in Figure 9.6.

9.1.5 Musician Spatialization

The musician spatialization features allows a local musician to experience the remote musicians' instruments as spatialized sound sources within his own, local space. In other words, the instrument of the local musician whose virtual location is to his left will appear to come through the left headphone, while the instrument of the musician whose virtual location is to his right will appear to come through the right headphones. The spatialization effect is determined by the local musician's body orientation, and changes accordingly. To formalize this feature, assume that M musicians located at virtual locations $\vec{p}_i \in \mathbb{R}, i = 1, \ldots, M$ are each producing the source audio signals $s_i(t)$. Each musician i sets a base volume of v_i^{base} and receives an overall mix $\vec{m}_i = (m_{Li}, m_{Ri})$, representing the left and right audio channels. As

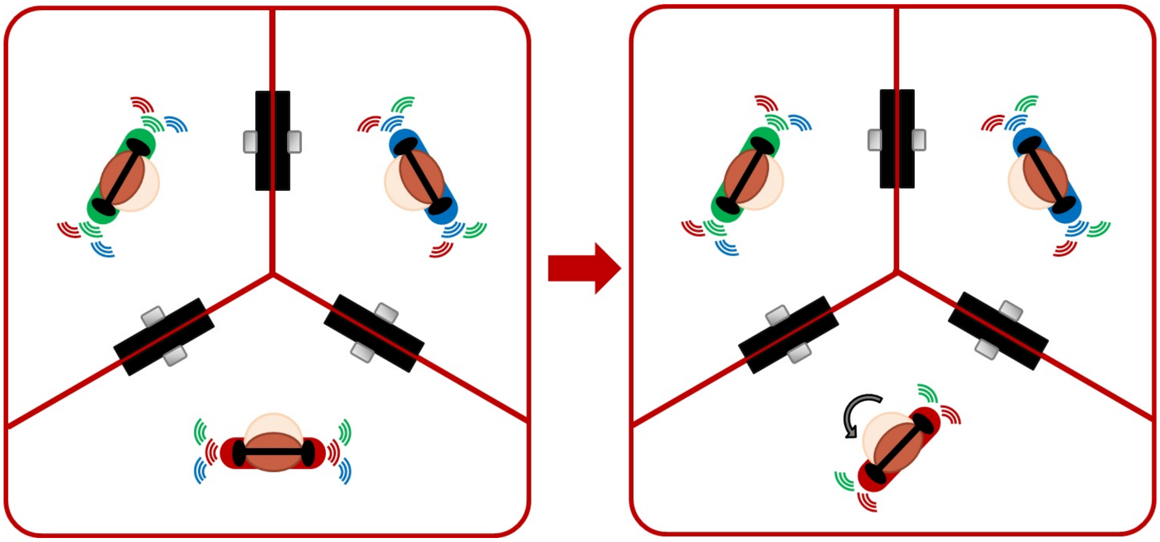

Fig. 9.6: Graphical respresentation of the track panning feature.

such, all musician receives an overall mix

$$m_k(t) = \sum_{j=1}^{M} \frac{v_j^{base}}{2} s_j(t) \tag{9.16}$$

where $k \in \{L\text{=left}, R\text{=right}\}$. When musician spatialization is in use, each musician i receives a mix

$$m_{ki}(t) = \frac{v_i^{base}}{2} s_i(t) + \sum_{j \neq i}^{M} a_{kij} \frac{v_j^{base}}{2} s_j(t) \tag{9.17}$$

where a_{kij} is a channel-wise mixing coefficient within the range between 0 and 1 that depends both on the distance between musicians i and j, and the orientation of musician i. We introduce the unit length vector $\vec{e_i}$, which points from the right to the left ear. We set

$$a_{kij} = \left(1 + b_k \frac{(\vec{p}_i - \vec{p}_j)}{\|\vec{p}_i - \vec{p}_j\|} \cdot \vec{e}_i\right)^2, \tag{9.18}$$

with $b_L = -1, \quad b_R = 1$. The scalar product between the difference vector and the ear-connection vector is within the range $[-1, 1]$, and the squaring ensures that the overall energy of the source signal remains constant when orienting the head towards a musician.

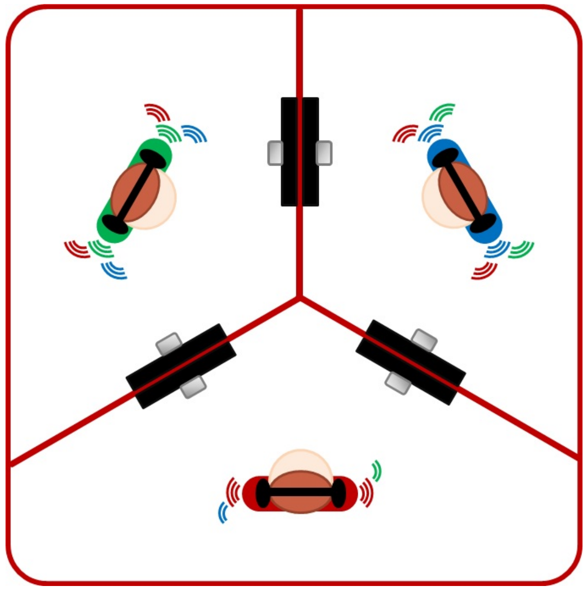

Fig. 9.7: Graphical representation of the musician spatialization feature.

We note that since both track panning and musician spatialization rely on body orientation, the two features cannot be used simultaneously. To prevent any errors, our system has been designed to automatically disable one of the features whenever the user chooses to enable the other.

9.2 Graphical User Interface

As described earlier, our responsive environment supplements shared video with a simple graphical user interface that not only give the musicians complete control over the system features, but also provides simple yet effective dynamic visual representations of the state of their performance at a glance. Through our long-term deployment with the three-piece band, and our participatory design cycle with the composer, we were able to further simply our responsive environment's GUI, which was also updated reflect the latest version of our system.

First, the number of introductory screens was reduced from two to one, in order to help shorten the start-up process. When musicians launch our SuperCollider software, they are now greeted with the introductory screen seen in Figure 9.8, where they can enter the IP addresses of those wishing to participate in the distributed performance. The system retains previous settings, allowing musicians to use the same IP addresses from an earlier performance, rather than having to enter them manually each time. The introductory screen also allows musicians to select the features they would like to use during the performance, independently from one another.

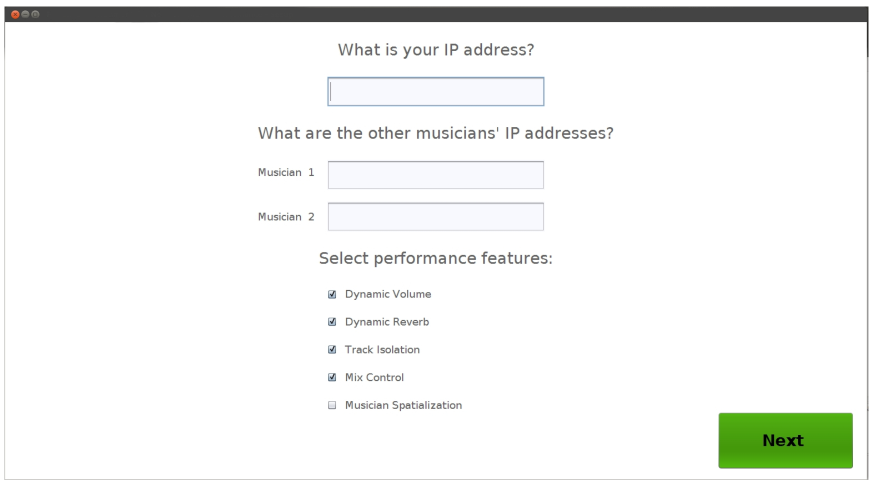

Fig. 9.8: Release candidate graphical user interface, introductory screen.

Once the parameters have been selected on the introductory screen, musicians can press the 'Next' button to move to the main screen, seen in Figure 9.9, which offers the following functionality:

- **Calibrating the Kinect:** Calibration only needs to be performed the first time a musician uses the system, or if the Kinect has been moved. After pressing the button, all a musician has to do is move around his space, making sure to cover the extremities of the horizontal plane seen by the Kinect. The latter can be determined by looking at the Kinect depth image viewer made available to users during calibration, as seen in Figure 9.10. A fixed number of sample points are collected, and used to determine the size of the local

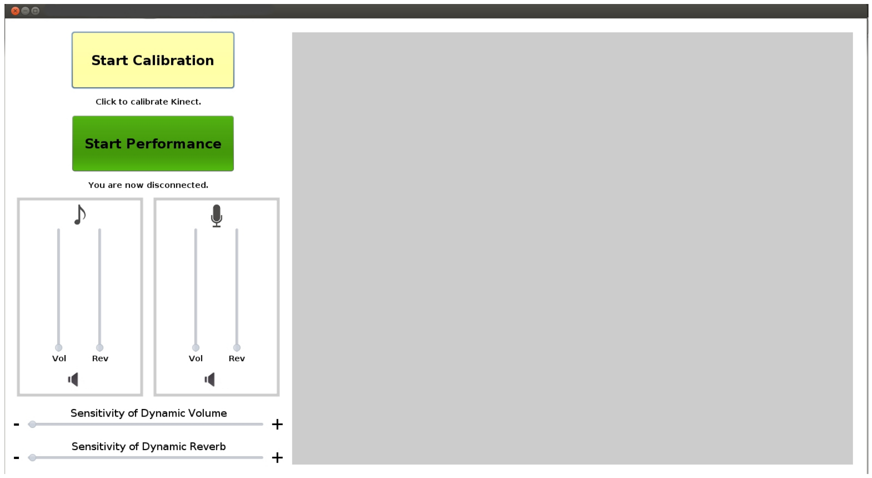

Fig. 9.9: Release candidate graphical user interface, main screen before the start of the performance.

musician's surroundings. These parameters are then used to map all locales into the shared space and, in turn, calculate the musicians' virtual locations relative to one another. During the calibration process, the musician receives progress information, and all other functions are disabled until completion.

- **Starting and Stopping the performance:** The alpha system required musicians to first "connect" to the performance, and prevented them from starting unless all participants had been connected, an undesirable side-effect our of software implementation at the time. The process was considerably simplified in the release candidate, which allows each local musician to join a performance simply by pressing the 'Start Performance' button. Doing so informs her if any other musicians have also joined the performance from their end, and enables her to immediately begin interacting with them. Remote musicians are assigned labels (e.g. 'Musician 1') that correspond with the IP address assignment completed earlier on the introductory screen. In the case where no remote participants are connected, the local musician can play her own instruments and adjust her settings(as described below) while waiting for others to join the performance. In addition, as seen in Figure 9.11, once pressed, the 'Start Per-

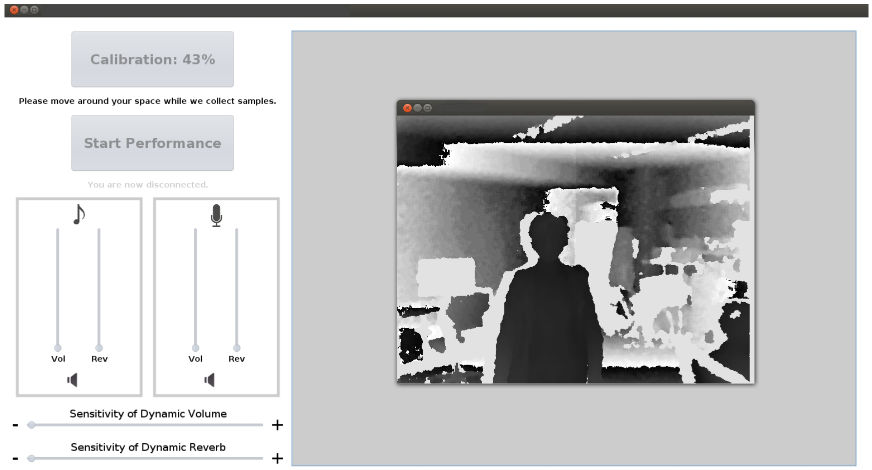

Fig. 9.10: Release candidate graphical user interface, main screen during calibration process.

formance' button turns into a 'Stop Performance' one, allowing the musician to go offline at the end of the performance.

- **Adjusting Settings:** Musicians can set the volume and reverb levels of their own instruments and microphones using the appropriate sliders. In addition, if any of the musicians choose to turn on the dynamic volume or dynamic reverb features, they can independently adjust these features' sensitivities using the corresponding sliders. After pressing the 'Start Performance' button, musicians are also presented with a smaller feature selection screen (seen in Figure 9.12) that allows them to independently change their selections mid-performance, should they wish to alter the settings previously chosen on the introductory screen.

- **Dynamic visual information:** Like its predecessor, the release candidate also provides musicians with dynamic graphical representations of their interactions with one another once they join a performance (see Figure 9.13). Such representations have been updated in accordance with the feedback received during the long-term deployment, and the participatory design cycle. First, the avatar's colours have been changed to better reinforce the metaphor of

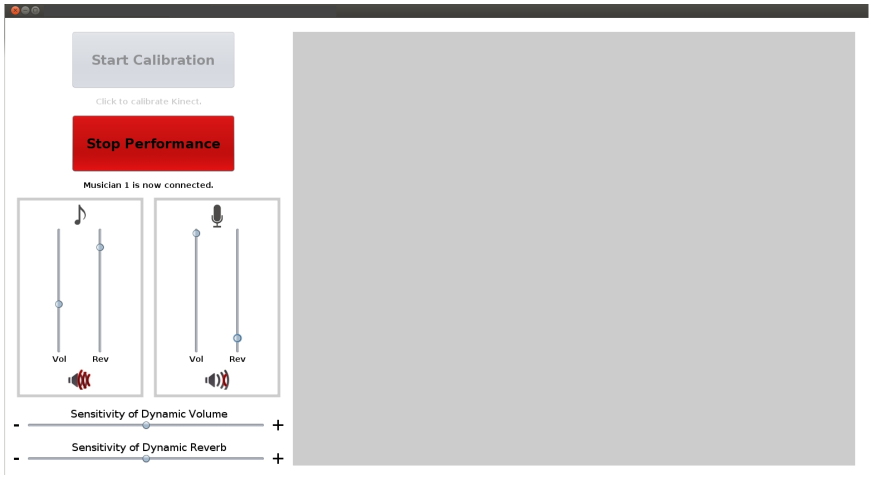

Fig. 9.11: Release candidate graphical user interface, main screen after the 'Start Performance' button is pressed.

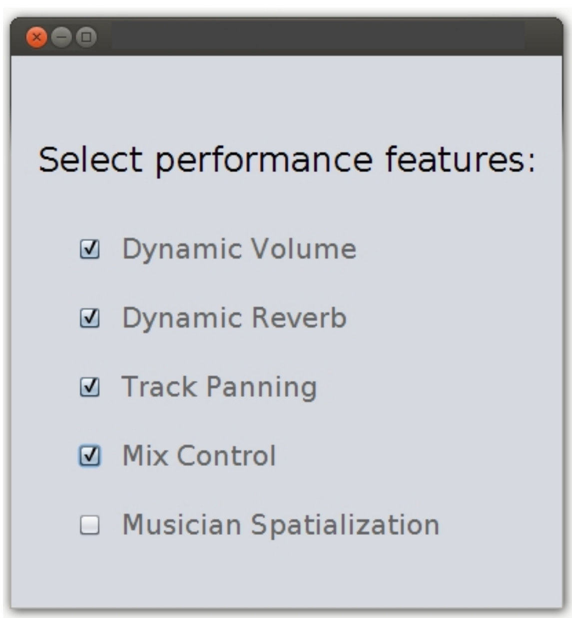

Fig. 9.12: Release candidate graphical user interface, feature selection screen.

bodies seen from a top-down view. In addition, the label for the local musician (previously set as ‘0’) has been removed, since each local musician can immediately infer that the middle avatar is hers based on our hardware configuration, which places a monitor on either side of her in a three-way collaboration. As per the composer’s suggestion, the sound waves were also assigned different colours (ranging from green to red) to help musicians dynamically gauge the upper limits of the changes induced by the dynamic volume features. Finally, black ‘echo’ waves were added to represent reverb levels. Since, to the best of our knowledge, no simple graphical representations of reverberation exist, we determined through consultation with Mr. Cowan that such waves best reflected the echoing nature of reverberation. A closeup of the updated dynamic animations can be seen in Figure 9.14

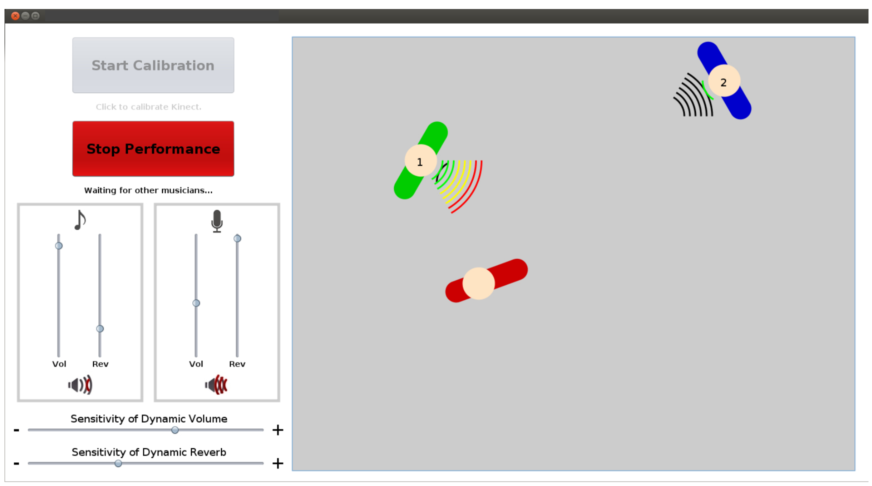

Fig. 9.13: Release candidate graphical user interface, main screen with graphical animations.

Finally, software for the release candidate described here has been made available to musicians on GitHub, at `http://www.github.com/delshimy/RENMP`. The minimum hardware required by each musician to try the system (albeit without shared video) is a computer and a Microsoft Kinect unit. Similarly, the responsive environment for musical composition, developed as part of our collaboration with Mr. Cowan,

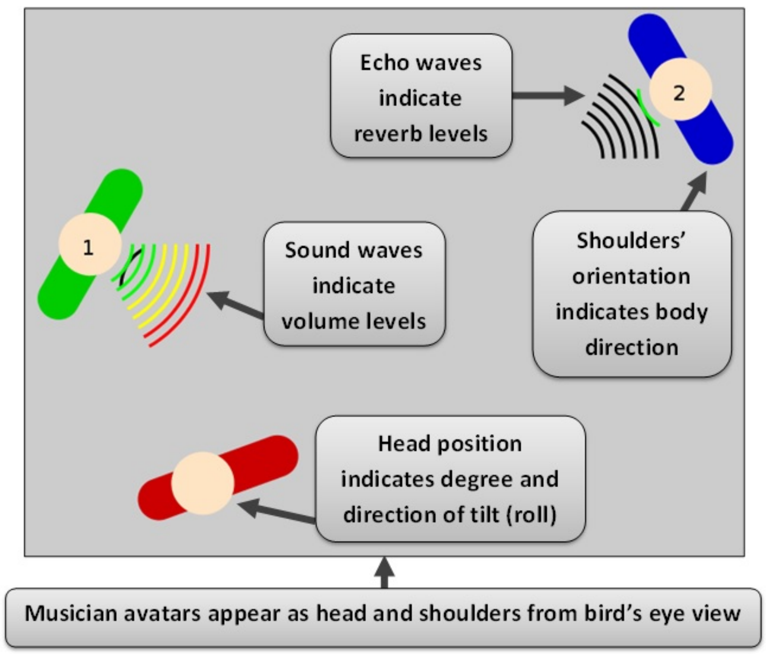

Fig. 9.14: A closeup of the graphical animations used as part of the release candidate's graphical user interface.

is also available for musicians to try at http://www.github.com/delshimy/REMC, and shares the minimum hardware requirement of a computer and Kinect Unit. Both software packages includes extensive installation and setup instructions, as well as detailed user manuals, ensuring that those wishing to try our artefacts can do so with relative ease.

Chapter 10

Lessons, Recommendations and Conclusions

While our efforts developing a responsive environment for distributed performance successfully produced two novel artefacts, we consider the lessons learned throughout the process to also be an important contribution of our work. At the onset of this project, one of our goals could haven described as investigating the extent to which traditional, iterative user-centered design principles, as typified by the works of Norman [143] and Gould and Lewis [81], could be applied to the systematic development of creative applications by designers who consider themselves non-artists. However, as described throughout this dissertation, our user-driven design of a novel interactive performance environment was met with a number challenges. In resolving these challenges, we were able to investigate how various human-computer interaction design techniques could be adapted to best fit our chosen application domain. We believe that the lessons drawn from our work could be of value to interface developers working with unique users, such as musicians, dancers, artists or actors, or on the support of creativity in various activities, such as performance, design, or problem-solving. Our suggestions are particularly aimed towards developers wishing to explore, understand or even augment artistic domains in which they would not consider themselves to be experts. This chapter outlines our key recommendations, along with suggestions for possible future work and our final conclusions.

10.1 Lessons and Recommendations

10.1.1 User Involvement at Various Stages of Design

New musical interfaces are typically created by their own users as a means of executing a specific vision, or conveying a well-defined message, the meaning of which can, in some cases, only be fully realized through user engagement. In contrast, our goal was not necessarily that of conveying a particular message, but rather that of providing tools that would allow the artists themselves to create and share their own message. While this could arguably be considered a design goal, deciding to facilitate an existing practice is a rather broad objective to tackle, especially when progress within such a practice may prove difficult to measure, as is the case with musical performance, and when specific use case scenarios are not particularly widespread, as is the case with distributed musical performance.

As a result, we opted to undertake every stage of the project from a strict user-driven perspective. However, somewhat against our initial expectations, we found that the level of user involvement exhibited a gradual but necessary shift throughout the project's life cycle. As described in Section 2.4, user-centered design involves a one-way exchange of information, from user to developer, whereas participatory design allows for a two-way exchange, back and forth between user and developer. Given that we adopted a user-centered design methodology during the early stages, followed by a long-term deployment and a participatory design approach towards the end, the role of the user changed from passive to increasingly active as development proceeded. Eventually, we even conceded our capacity as "experts" to the musicians themselves, and instead viewed our role as that of facilitators, developing technology to support the end users in executing their vision. Overall, maintaining such a fluidity between the responsibilities entailed by the roles of user and designer was crucial to the success of our efforts: it was only through the one-sided user observations, interviews and early testing that we were able to develop simple prototypes that, in turn, became crucial towards grounding our collaboration with the three-piece band during the long-term deployment, and the cooperative prototyping technique used in the participatory design phase. These latter two stages were particularly significant in shaping our release candidate in a manner we believe might not have been possible had we exclusively employed a user-centered design methodology. Furthermore, an important aspect of our role as facilitators was ensuring that user input was not driven by idiosyncratic needs, which can readily occur when working within an artistic context. In turn, this would allow any changes to the system to remain potentially appealing and beneficial to a larger proportion of the demographic constituting the target end user. This also necessitated that we negotiate clear boundaries such that user suggestions for new functionality were reasonably feasible within the constraints of available tools and technology.

10.1.2 Approaching the Design of a Novel System for Artistic or Creative Use

Applying a systematic procedure to the design of novel technology intended for creative or artistic use can be a difficult endeavour. First, user needs are not easily defined within such contexts. Thus, it is difficult to fully conceive of the system's functionality during the early stages of design. This, in turn, may complicate communication and collaboration with end users, especially if they have no prior experience with comparable systems. Furthermore, common evaluation techniques, while suitable for evaluating metrics that may be indicative of overall usability, may prove to be inadequate when exported to non-utilitarian domains. To resolve such issues, we outline three key recommendations when designing novel creative tools from a user-driven perspective.

- **Validate the basics:** Developers looking to support or augment creative or artistic activities may encounter difficulties defining concrete user needs at the start of a project, especially if the goal is to introduce technology that is completely novel to the typical end user. This does not necessarily imply that a system may not come to serve a purpose. It does, however, mean that designers must first develop a thorough understanding of how users would normally undertake the most basic aspects of the activities related in nature to the intended system. In turn, such an understanding can serve as the foundation for conceptualizing elementary system functionality, which in turn can become a starting point for development, as was the case with our experience.

Despite the variety in age, musical style, and musical aspirations, none of the musicians we worked with throughout the project's life cycle had previously participated in real-time distributed performance. This is not surprising, considering that typical Internet connections make online musical collaboration rather difficult. Therefore, we could not anticipate the types of interaction they would find useful or the problems they may encounter. We were able to overcome this challenge by defining our system functionality in stages throughout the project's lifecycle, rather than entirely at once when we first conceived of the idea of augmenting network musical performance. By starting out with the first incarnation of dynamic volume, an easily understood function that utilized simple motion in a co-present setting to affect one of the most basic aspects of an instrument's sound, we were able to rapidly prototype an environment exemplifying the basic notion of augmenting performance via performer-performer interaction. In turn, such a prototype helped lay the foundation for eliciting user feedback, and created a basis for subsequent discussions with target users, which had previously proven tricky to ground. From this point on, the system features increased in complexity only as a direct result of user input, ensuring that all controls remained simple and mappings transparent.

- **Investigate alternatives to "usability":** As described previously, one of our goals entailed investigating the applicability of standard user-centered techniques to the design of novel musical interfaces. Nonetheless, as early as our user observation stage, it became apparent that several aspects of traditional usability testing would not be applicable to the system we had in mind: evaluation metrics for usability typically encompass such criteria as task completion time and rate, error rate, accuracy and overall satisfaction [133] and, as a result, are closely tied to task-based design and evaluation. However, to what extent could we suitably define a user task within musical performance, particularly when the focus is on performer-performer rather than performer-instrument interactions? And if musicians did not strictly *need* to take advantage of our system's features, how could we evaluate their level of satisfaction? Similar sentiments have, for some time, been echoed by a subset of the HCI community, as demonstrated by the emergence of a "third-wave", or experience-based approach to design that could adequately address both the hedonic and pragmatic aspects of interaction. Nonetheless, the question remains of how to determine which of these aspects are particularly relevant within the context of a specific activity.

In our case, such challenges were exacerbated by the fact that our background and knowledge as developers did not match those of our target end users. As a result, we sought to investigate, in as systematic a fashion as possible, which evaluation criteria would specifically fit musicians and their expectations. This was accomplished, as described in Section 5.3, via early user interviews that uncovered such benchmarks as enjoyment, creativity and self-expression as being of utmost importance to musicians. Although such criteria may seem obvious in retrospect, as non-musicians, we might easily have devised an entirely different list. Naturally, such a list would have, at best, been based purely on assumptions. Therefore, we posit that extracting benchmarks directly from the users themselves, by means of observations, discussions or interviews, can be far more effective in producing accurate results.

As a result, we encourage developers interested in taking a user-driven approach to design of non-utilitarian system to investigate alternatives to the traditional notion of "usability" by uncovering and defining benchmarks specifically suited to the activity at hand. In turn, such benchmarks can lead to a more reliable evaluation of the system against the target user's expectations.

- **Tailor evaluation techniques to fit the nature of the project:** According to MacDonald and Atwood, "[e]valuation has been a dominant theme in

HCI for decades, but it is far from being a solved problem" [133]. The authors further describe a lack of frameworks when it comes to evaluating the user experience in a holistic manner that encompasses both the hedonic and the pragmatic aspects of interaction. This is particularly evident in non-utilitarian application domains: as described in Section 2.5, developers of creative or artistic systems who find existing quantitative methods unsuitable for their needs often resort to devising their own techniques. While this has contributed to a lack of standards for design and evaluation, which has been criticized by some within the context of music-oriented HCI, we argue that this, perhaps, may simply be an unavoidable phenomenon: the user experience with playful or creative interfaces is often marked by an idiosyncratic quality. As such, while utilitarian systems can successfully be tested according to established standards, perhaps the search for a "one size fits all" solution for third-wave HCI may be considered futile. Instead, we propose that developers investigate the possibility of *adapting* existing techniques such as those detailed in Section 2.5 to their needs, an approach also advocated by Kiefer et al. [112], or devising new ones if necessary. Selection from existing techniques may, in turn, be motivated by several factors, including the availability of necessary tools, the degree to which a technique's intended context matches the one under examination, and the level of modification required to adapt a technique from one application to another.

To give an example from our efforts, after defining our evaluation criteria such as enjoyment, interaction with others, creativity and self-expression, we found that, to the best of our knowledge, standardized methods for assessing them had yet to be established. IJsselsteijn's gaming experience questionnaire [97], which largely encompasses questions on flow and immersion, proved a suitable contender for the evaluation of enjoyment due to the breadth of behaviours it examined. The GEQ was also chosen due to its inclusion of a component on social presence, which, when supplemented with logged position data, helped shed further light on the musicians' level of interaction with one another. Furthermore, the general nature of the GEQ's questions meant that relatively little modification was necessary to adapt it to the musical context. The GEQ's enduring popularity in ludology also inspired us to devise our own questionnaires based on the behaviours and values typically associated with self-expression and creativity.

As another example, when it came time to transform our alpha system into its next iteration, we decided against traditional, formal user tests with multiple users, and opted instead to design a long-term deployment cycle with a fixed set of users. The mixed research techniques used during this cycle allowed us to supplement our quantitative questionnaires and logged data with qualitative insight acquired via user discussions, thereby increasing the depth of feedback we could elicit from our users. In turn, the quality of insight that were able to gain from our target users contributed directly towards improving their experience with our responsive environment.

10.1.3 Augmenting Computer-Supported Cooperative Work

As described at the beginning of this dissertation, our objectives with respect to Computer-Supported Cooperative Work were to 1) support the creative, ludic and spontaneous aspects of social interaction within a distributed context, and 2) explore whether distributed collaboration could improve on its co-present counterpart by leveraging its underlying technology towards further assisting target users in effectively accomplishing the task at hand. We sought to establish the former by adhering to the user-centric philosophy advocated in the context of CSCW by Ackerman [1], Rodden and Blair [161], and Cartensen and Schmidt [42], and that of musical performance by Fencott and Bryan-Kinns [71] and Healey et al. [93], all of whom regard successful collaboration as contingent on a thorough understanding of target participants and their collaborative behaviours. In our case, this entailed observing the interpersonal interactions of co-located performers, and subsequently eliciting the specificities of these interactions by means of interviews. Having established creativity, enjoyment, self-expression and social interaction as the performance criteria most valued by musicians, we consequently sought to understand and improve the ways in which we could better support such benchmarks in a distributed setting.

More importantly, however, we strove to capitalize on the knowledge we acquired from our users during the early stages of the project to better address our second objective. Namely, we wanted to examine how the shortcomings of distributed collaboration could be resolved in a manner that not only bridges the gap between the co-present and distributed contexts, but also serves to further enhance those aspects of the activity at hand that our users had deemed most valuable. The underlying basis of telepresence research, the most prominent example of the "same time/different place" category of CSCW, has been to engender, as best as possible, a feeling of co-presence through the support of the non-verbal cues and gestures that are typically poorly conveyed between remote participants. The problem with such an approach is that it can only, at best, mimic co-location. Within the context of network performance, a parallel methodology is perhaps best illustrated through the breadth of research that aims to decrease latency and increase bandwidth as a means of facilitating musical collaboration. We argue, however, that the goal of distributed systems should not stop at simply mirroring their co-present counterparts. Distributed collaborative environments must, by nature, introduce a certain level of technology to offer their users support over even the most basic aspects of cooperation. As a result, we question whether participants stand to benefit from developers leveraging the technology at their disposal towards *augmenting* the activities these systems afford.

In our case, supporting audio sharing, the most elementary aspect of network performance, meant equipping each location with tools that musicians do not require under normal circumstances. With the addition of a Microsoft Kinect, we were able capitalize on the computing power necessary to make distributed performance a possibility, in a bid to present the network as a unique and appealing medium in its own right to the less technologically inclined musicians. In the end, our goal of augmenting an existing activity in a manner that utilizes existing, well-understood embodied interactions helped the musicians perceive distributed performance as a activity in which they would likely partake again, as expressed by members of the three-piece ensemble at the end of our long-term collaboration.

Furthermore, our work with the composer indicated that our system features were useful to aspects of performance that extended beyond the relatively uncommon distributed context, and into the more widespread scenarios of composition, mixing recording. Thus, by exploring a specific activity under more challenging circumstances, we were able to gain insight on how it may also be improved under less demanding conditions. As a result, we suggest that telepresence research could similarly benefit from an approach that seeks to augment distributed collaboration, rather than simply echo its co-present counterpart.

10.2 Open Issues and Future Work

10.2.1 Further Investigation of Latency

While our hardware configuration ensured a latency well below the established ensemble performance threshold, we acknowledge that the average musician, who typically has access only to a standard Internet connection, will experience different results. Our latency session, held as part of the long-term deployment, was intended to give us an idea of how well the system features would hold up under increasing latencies that eventually placed our system into the realm of the Latency-Accepting Approach. As described earlier, the musicians struggled to maintain a rhythm once the latency reached or surpassed 100 ms, but nevertheless continued to use the system features. A question worth investigating is whether the system features could themselves be adjusted to better accommodate greater levels of latencies, or whether entirely new features could be designed to address that issue. For instance, Schroeder and Rebelo have indicated that different rhythmic landscapes were better suited to the nature of the network and its inherent delays. As such, we question whether system features that might encourage musicians to explore new rhythmic patterns would help them not only adapt, but also further explore, the network itself as "a space for being" [167].

10.2.2 Additional System Features

In addition to features that may help musicians adapt to greater levels of latency, we believe that our responsive environment could benefit from functionality of a more artistic or inventive nature. As described above, an important aspect of our role as developers throughout the long-term deployment and participatory design cycle was to "reign in" user suggestions to ensure that they maintained a greater level of general applicability than idiosyncrasy. However, given that our current system has proven successful in supporting many practical aspects of mixing and performance, we now have an opportunity to begin exploring more creative options, such as the keyboardist's suggestion for psychedelic video. We believe that this would be best achieved through another participatory design cycle, as such features would largely evolve according to an artist's vision. These new additions may not necessarily hold universal appeal, but rather serve to support artistic expression. As such, we would need to explore and develop them on a case by case basis, creating customized responsive environments for composers or ensembles wishing to explore specific effects through embodied interaction.

10.2.3 Quantitative Studies with Release Candidate

The long-term deployment and participatory design cycles have allowed us to tailor our system to meet musicians' expectations, not only in terms of functionality, but in overall quality of sound and interaction. As both the long-term deployment and participatory design were held somewhat concurrently, the composer was able to test and approve the band's suggestions, and vice-versa. However, while we tried to steer the musicians' feedback towards improvements with a broad appeal, we posit that our system could benefit from additional validation by means of standard quantitative experiments with a greater number of users. Not only would this allow us to gain additional insight into the system, but it would help shed further light on the validity of our design process itself, and our decision to collaborate closely with a small number of users, rather than test broadly with a larger population.

10.3 Conclusion

A responsive environment for distributed performance, was designed to augment collaboration between displaced musicians. Through the use of five unique features, dynamic volume, dynamic reverb, track panning, mix control and musician spatialization, musicians are able to seamlessly create and alter individualized mixes mid-performance, simply by moving around their space. By responding to changes in position and orientation, our system allows musicians to utilize its features without having to detach themselves from the primary task of music-making. Our responsive environment also supplements shared video with dynamic visual representations of the environment via simple graphical animations. To the best of our knowledge, our solution is the only distributed performance system of its kind that simultaneously 1) exports the notion of "shared space" from the CSCW domain to the distributed performance context, allowing musicians to perceive local and remote environments as simple extensions of one another, 2) uses shared space as a means to restore the spatialization of musical instruments that is inherent to the co-present context, yet lost in the distributed one, 3) capitalizes on embodied interactions as a means of control, and 4) offers performers the ability to affect one another's sound parameters through their interpersonal interactions. Together, such properties allow our responsive environment to confer a greater level of co-presence than traditional solutions for online performance.

Employing a number of user-driven techniques, our responsive environment has evolved through several stages to reach its current incarnation. Following the mandates of standard user-centered design, we began with a thorough focus on the user by means of observations, interviews and persona profiles. This information inspired a number of early prototypes that, in turn, were used to elicit user feedback via formal tests. Subsequently, we increased the level of user participation, first through a long-term deployment and collaboration with a three-piece ensemble, incorporating their feedback and suggestions at every stage. Questionnaires collected throughout the collaboration proved that our system helped enhance the musicians' sense of enjoyment and self-expression. Furthermore, qualitative analysis of our discussions with the band members has shown that they found the system practical, and would likely use it again. In addition, a participatory design cycle was held with a composer who was asked to write a few pieces using our system in order to evaluate its creative potential, and make suggestions for improvement. While the composer was aware of the system's intended use case scenario of distributed performance, he nevertheless found the system features to be useful within the context of mixing and recording. Thus, the result of the participatory design not only included, as expected, a series of refinements and additions to our responsive environment, but also an altogether new artefact: a responsive environment for musical composition.

Our efforts have enabled us to explore a unique and demanding form of computer-mediated communication. Rather than attempt to imitate the co-present scenario, as many telepresence or videoconferencing systems do, we chose instead to leverage the technology inherent to the distributed context towards meaningfully augmenting the activity at hand. The result was a system that helped musicians regard network musical performance as not only a viable option, but one that is appealing and exciting in its own right. Taking our inspiration from user-driven research in Computer-Supported Cooperative Work, we placed the vast majority of our focus on exploring the subtleties and complexities of our users' collaborative behaviours, rather than the system's infrastructure. By seeking to understand musicians' needs under an extreme condition such as network performance, we were able to support these needs in a manner that extended their intended context: as the participatory design cycle illustrated, the system features we designed with help from end users proved to be both useful and suitable for artistic expression during composition and mixing, a far more common and widespread scenario than distributed performance.

Our choice of user-driven approach also necessitated that we tailor all design benchmarks and evaluation techniques to the exacting nature of musical performance. However, the challenges we encountered in terms of user understanding and system evaluation are by no means unique to the context of musical performance, but inherent to many creative and artistic domains. Therefore, we believe that our approach to resolving those challenges can be of use to interface developers looking to acquire a deeper understanding of the user experience that the traditional notion of usability alone does not necessarily afford.

References

[1] Mark S. Ackerman. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human-Computer Interaction*, 15(2):179–203, September 2000.

[2] Sigurdur O. Adalgeirsson and Cynthia Breazeal. MeBot: A Robotic Platform for Socially Embodied Presence. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, HRI '10, pages 15–22, Piscataway, NJ, USA, 2010. IEEE Press.

[3] Matthew Aldrich, Akash Badshah, Brian Mayton, Nan Zhao, and Joseph A. Paradiso. Random Walk and Lighting Control. In *Sensors, 2013 IEEE*, pages 1–4, Nov 2013.

[4] Chrisoula Alexandraki and Demosthenes Akoumianakis. Exploring New Perspectives in Network Music Performance: The Diamouses Framework. *Computer Music Journal*, 34(2):66–83, Jun 2010.

[5] Àlvaro Barbosa. Displaced Soundscapes: A Survey of Network Systems for Music and Sonic Art Creation. *Leonardo Music Journal*, 13, 2003.

[6] Àlvaro Barbosa and Martin Kaltenbrunner. Public Sound Objects: A Shared Musical Space on the Web. In *Proceedings of International Conference on Web Delivering of Music (WEDELMUSIC 2002)*, pages 9–15. IEEE Computer Society Press, 2002.

[7] Alissa N. Antle, Greg Corness, and Allen Bevans. Springboard: Designing Image Schema Based Embodied Interaction for an Abstract Domain. In David England, editor, *Whole Body Interaction*, Human-Computer Interaction Series, pages 7–18. Springer London, 2011.

[8] Alissa N. Antle, Greg Corness, and Milena Droumeva. Human-Computer-Intuition? Exploring the Cognitive Basis for Intuition in Embodied Interaction. *International Journal of Arts and Technology*, 2(3):235–254, 2009.

[9] Ronald M. Baecker, Jonathan Grudin, William Buxton, and Saul Greenberg. *Readings in Human-Computer Interaction: Toward the Year 2000*. Morgan Kaufmann, January 1995.

[10] Saskia Bakker, Alissa N. Antle, and Elise Van Den Hoven. Embodied Metaphors in Tangible Interaction Design. *Personal Ubiquitous Computing*, 16(4):433–449, April 2012.

[11] Liam J. Bannon. CSCW: An Initial Exploration. *Scandinavian Journal of Information Systems*, 5:3–24, August 1993.

[12] Liam J. Bannon. A Human-Centred Perspective on Interaction Design. In Antti Pirhonen, Pertti Saariluoma, Hannakaisa Isomki, and Chris Roast, editors, *Future Interaction Design*, pages 31–51. Springer London, 2005.

[13] Liam J. Bannon and Kjeld Schmidt. CSCW: Four Characters in Search of a Context. In John M. Bowers and Steven D. Benford, editors, *Studies in Computer Supported Cooperative Work*, pages 3–16. North-Holland Publishing Co., Amsterdam, The Netherlands, 1991.

[14] Olivier Bau, Atau Tanaka, and Wendy E Mackay. The A20: Musical Metaphors for Interface Design. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 8 of *NIME '08*, pages 91–96, 2008.

[15] Frank Bentley, Konrad Tollmar, David Demirdjian, Kimberle Koile, and Trevor Darrell. Perceptive Presence. *IEEE Computer Graphics Applications*, 23(5):26–36, September 2003.

[16] Zafer Bilda, Chris Bowman, and Ernest Edmonds. Experience Evaluation of Interactive Art: Study of GEO Landscapes. In *Proceedings of the 5th Australasian Conference on Interactive Entertainment*, IE '08, pages 1:1–1:10, New York, NY, USA, 2008. ACM.

[17] Zafer Bilda, Ernest Edmonds, and Linda Candy. Designing for Creative Engagement. *Design Studies*, 29(6):525–540, November 2008.

[18] Tina Blaine and Sidney Fels. Collaborative Musical Experiences for Novices. *Journal of New Music Research*, 32(4):411–428, 2003.

[19] Tina Blaine and Sidney Fels. Contexts of Collaborative Musical Experiences. In *NIME'03*, pages 129–134, 2003.

[20] Sara A. Bly, Steve R. Harrison, and Susan Irwin. Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. *Communications of the ACM*, 36(1):28–46, January 1993.

[21] Bert Bongers. Interaction in Multimedia Art. *Knowledge-Based Systems*, 13(7-8):479–485, 2000.

[22] Bert Bongers. Physical Interfaces in the Electronic Arts: Interaction Theory and Interfacing Techniques for Real-time Performance. In *Trends in Gestural Control of Music*, pages 41–70, 2000.

[23] Bert Bongers. Interactivating Spaces. In *Symposium on Systems Research in the Arts at the 14th Annual Conference on Systems Research, Informatics and Cybernetics*, August 2002.

[24] Bert Bongers. Interaction with the Disappeared Computer. In *Physical Interaction Workshop on Real World Interfaces at the Mobile HCI Conference*, September 2003.

[25] Jonas Braasch. Tintinnabulate and SoundWIRE: Tele-Colonization. `http://symphony.arch.rpi.edu/~braasj/TeleColoniztion_LinerNotes.pdf`, 2007. Accessed April 8, 2014.

[26] Jonas Braasch, Daniel L. Valente, and Nils Peters. Sharing Acoustic Spaces over Telepresence Using Virtual Microphone Control. In *Audio Engineering Society Convention 123*, October 2007.

[27] Jonas Braasch, Doug Van Nort, Pauline Oliveros, and Ted Krueger. Telehaptic Interfaces for Interpersonal Communication within a Music Ensemble. In *Proceedings of Meetings on Acoustics*, volume 19, June 2013.

[28] Nick Bryan-Kinns, Robin Fencott, Oussama Metatla, Shahin Nabavian, and Jennifer G. Sheridan. Interactional Sound and Music : Listening to CSCW, Sonification, and Sound Art. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '10, pages 403–406, 2010.

[29] Nolan Bushnell. Relationships Between Fun and the Computer Business. *Communications of the ACM*, 39(8):31–37, August 1996.

[30] Bill Buxton. Chunking and Phrasing and the Design of Human-Computer Dialogues. In *Proceedings of the IFIP World Computer Congress*, pages 475–480. North Holland Publishers, 1986.

[31] Bill Buxton. Ontario Telepresence Project: Final Report. In *Encyclopedia of Library and Information Sciences*. Information Technology Research Centre, Telecommunications Research Institute of Ontario, 1995.

[32] Bill Buxton. Living in Augmented Reality: Ubiquitous Media and Reactive Environments. In *Video Mediated Communication*, pages 363–384. Erlbaum, 1997.

[33] Linda Candy, Shigeki Amitani, and Zafer Bilda. Practice-led Strategies for Interactive Art Research. *CoDesign*, 2(4):209–223, December 2006.

[34] Linda Candy and Koichi Hori. The Digital Muse: HCI in Support of Creativity. *Interactions*, 10(4):44–54, July 2003.

[35] Gail Cantor. eJAMMING AUDiiO. http://www.ejamming.com. Accessed July 21, 2011.

[36] Brad Cariou. Design of an Alternate Controller from an Industrial Design Perspective. In *Proceedings of the 1992 International Computer Music Conference*, pages 366–367, 1992.

[37] Alexander Carôt. Soundjack. http://www.soundjack.eu/. Accessed July 21, 2011.

[38] Alexander Carôt, Ulrich Krämer, and Gerald Schuller. Network Music Performance (NMP) in Narrow Band Networks. In *Audio Engineering Society Convention 120*, May 2006.

[39] Alexander Carôt, Pedro Rebelo, and Alain B. Renaud. Networked Music Performance: State of the Art. In *Audio Engineering Society Conference 30th International Conference: Intelligent Audio Environments*, March 2007.

[40] Alexander Carôt and Christian Werner Fischinger. *Towards a Comprehensive Cognitive Analysis of Delay-Influenced Rhythmical Interaction*, pages 473–476. MPublishing, University of Michigan Library, 2009.

[41] Juan P. Carrascal and Sergi Jordà. Multitouch Interface for Audio Mixing. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '11, pages 100–103, 2011.

[42] Peter H. Carstensen and Kjeld Schmidt. Computer Supported Cooperative Work: New Challenges to Systems Design. In Kenji Itoh, editor, *Handbook of Human Factors*, pages 619–636, 1999.

[43] Chris Chafe. Ping: Interactive Music from Network Pings. https://ccrma.stanford.edu/~cc/shtml/2001ping.shtml. Accessed June 18, 2011.

[44] Chris Chafe, Scott Wilson, Al Leistikow, Dave Chisholm, and Gary Scavone. A Simplified Approach to High Quality Music and Sound over IP. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, pages 159–164, 2000.

[45] Noel Chateau and Marc Merisol. AMUSE: A Tool for Evaluating Affective Interfaces. In *CHI'05 Workshop on Evaluating Affective Interfaces*, 2005.

[46] Wei-Chao Chen, Herman Towles, Lars Nyland, Greg Welch, and Henry Fuchs. Toward a Compelling Sensation of Telepresence: Demonstrating a Portal to a Distant (Static) Office. In *Proceedings of the Conference on Visualization '00*, VIS '00, pages 327–333, Los Alamitos, CA, USA, 2000. IEEE Computer Society Press.

[47] Elaine Chew, Alexander Sawchuk, Carley Tanoue, and Roger Zimmermann. Segmental Tempo Analysis of Performances in Performer-Centered Experiments in the Distributed Immersive Performance Project. In *Proceedings of International Conference on Sound and Music Computing '05 (SMC05)*, November 2005.

[48] Jeremy R. Cooperstock. History of Spatially Distributed Performance. `http://www.cim.mcgill.ca/sre/projects/rtnm/history.html`. Accessed July 28, 2011.

[49] Jeremy R. Cooperstock. McGill Ultra-Videoconferencing. `http://www.cim.mcgill.ca/sre/projects/rtnm/`. Accessed July 28, 2011.

[50] Jeremy R. Cooperstock. From the flashing 12:00 to a usable machine: applying ubicomp to the vcr. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, pages 281–282, New York, NY, USA, 1997. ACM.

[51] Jeremy R. Cooperstock, Sidney S. Fels, William Buxton, and Kenneth C. Smith. Reactive Environments: Throwing Away your Keyboard and Mouse. *Communications of the ACM*, 40:65–73, 1997.

[52] Jeremy R. Cooperstock and Stephen P. Spackman. The Recording Studio that Spanned a Continent. In *Proceedings of the International Conference on Web Delivering of Music (WEDELMUSIC 2001)*, pages 161–167, Washington, DC, USA, 2001. IEEE Computer Society.

[53] Jeremy R. Cooperstock, Koichiro Tanikoshi, Garry Beirne, Tracy Narine, and William A. S. Buxton. Evolution of a Reactive Environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 170–177, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[54] Caroline J. Cornelius, Mai Anh Nguyen, Caroline C. Hayes, and Roopa Makena. Supporting Virtual Collaboration in Spatial Design Tasks: Are Surrogate or Natural Gestures More Effective? *IEEE T. Human-Machine Systems*, 43(1):92–101, 2013.

[55] Greg Corness and Thecla Schiphorst. Performing with a System's Intention: Embodied Cues in Performer-system Interaction. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*, C&C '13, pages 156–164, New York, NY, USA, 2013. ACM.

[56] Mihaly Csíkszentmihályi. *Flow: The Psychology of Optimal Experience*. Harper Perennial, New York, NY, March 1991.

[57] Pieter M. A. Desmet, Paul Hekkert, and J. J. Jacobs. When a Car Makes you Smile: Development and Application of an Instrument to Measure Product Emotions. *Advances in Consumer Research*, 27:111–117, 2000.

[58] Steve Dixon. *Digital Performance: A History of New Media in Theatre, Dance, Performance Art, and Installation*, chapter 17. The MIT Press, Cambridge, MA, 2007.

[59] Paul Dourish. *Where the Action is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, MA, USA, 2001.

[60] Paul Dourish. Embodied Interaction: Exploring the Foundations of a New Approach to HCI. In J.M. Carroll, editor, *Human-Computer Interaction in the New Millennium*, ACM Press Series. ACM Press, 2002.

[61] Paul Dourish and Sara Bly. Portholes: Supporting Awareness in a Distributed Work Group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 541–547, New York, NY, USA, 1992. ACM.

[62] William Duckworth. Cathedral. `http://www.monroestreet.com/Cathedral/main.html`. Accessed July 22, 2011.

[63] Ernest Edmonds. Abstraction and Interaction: An Art System for White Noise. *International Conference on Computer Graphics, Imaging and Visualization*, 0:423–427, 2006.

[64] Peter Eisert. Immersive 3-D Video Conferencing: Challenges, Concepts, and Implementations. *Proceedings of the International Society for Optical Engineering*, pages 69–79, 2003.

[65] Scott Elrod, Gene Hall, Rick Costanza, Michael Dixon, and Jim Des Rivières. Responsive Office Environments. *Commun. ACM*, 36:84–85, July 1993.

[66] Daniel Fallman and John Waterworth. Dealing with User Experience and Affective Evaluation in HCI Design : A Repertory Grid Approach. In *Proceedings of CHI 2005, Conference on Human Factors in Computing Systems*, pages 2–7, 2005.

[67] Mark Christopher Feldmeier. *Large Group Musical Interaction Using Disposable Wireless Motion Sensors*. PhD thesis, Massachusetts Institute of Technology, 2002.

[68] Sidney Fels. Designing for Intimacy: Creating New Interfaces for Musical Expression. *Proceedings of the IEEE*, 92(4):672–685, Apr 2004.

[69] Sidney Fels and Kenji Mase. Iamascope: A Graphical Musical Instrument. *Computers and Graphics*, 2:277–286, 1999.

[70] Robin Fencott. *Computer Musicking*. PhD thesis, Queen Mary, University of London, 2012.

[71] Robin Fencott and Nick Bryan-Kinns. Computer Musicking: HCI, CSCW and Collaborative Digital Musical Interaction. In Simon Holland, Katie Wilkie, Paul Mulholland, and Allan Seago, editors, *Music and Human-Computer Interaction*, Springer Series on Cultural Computing, pages 189–205. Springer London, 2013.

[72] Owen Noel Newton Fernando, Kazuya Adachi, Uresh Duminduwardena, Makoto Kawaguchi, and Michael Cohen. Audio Narrowcasting and Privacy for Multipresent Avatars on Workstations and Mobile Phones. *IEICE - Transactions on Information and Systems*, E89-D:73–87, January 2006.

[73] Jon Forsyth, Aron Glennon, and Juan P. Bello. Random Access Remixing on the iPad. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '11, pages 487–490, 2011.

[74] R. A. Fowles. Symmetry in Design Participation in the Built Environment: Experiences and Insights from Education and Practice. In *Proceedings of CoDesigning*, pages 59–70. Springer, 2000.

[75] Justin Frankel, Christophe Thibault, and John Schwartz. Ninjam. `http://www.ninjam.com`. Accessed June 18, 2011.

[76] Christian Geiger, Holger Reckter, David Paschke, Florian Schutz, and Cornelius Poepel. Towards Participatory Design and Evaluation of Theremin-based Musical Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '08, pages 303–306, Genova, Italy, June 2008.

[77] Donald R. Gentner and Jonathan Grudin. Why Good Engineers (Sometimes) Create Bad Interfaces. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, CHI '90, pages 277–282, New York, NY, USA, 1990.

[78] Simon J. Gibbs, Constantin Arapis, and Christian J. Breiteneder. TELEPORT: Towards Immersive Copresence. *Multimedia Systems*, 7(3):214–221, May 1999.

[79] Rolf Inge Godøy, Alexander Refsum Jensenius, and Kristian Nymoen. Chunking in Music by Coarticulation. *Acta Acustica united with Acustica*, 96(4):690–700, 2010.

[80] John D. Gould, Stephen J. Boies, Stephen Levy, John T. Richards, and Jim Schoonard. The 1984 Olympic Message System: A Test of Behavioral Principles of System Design. *Communications of the ACM*, 30:758–769, September 1987.

[81] John D. Gould and Clayton Lewis. Designing for Usability: Key Principles and What Designers Think. *Communications of the ACM*, 28:300–311, March 1985.

[82] Scot Gresham-Lancaster. The Aesthetics and History of the Hub: The Effect of Changing Technology on Network Computer Music. *Leonardo Music Journal*, 8:39–44, 1998.

[83] Mick Grierson. Noisescape: An Interactive 3D Audiovisual Multi-User Composition Environment. In *Proceedings of International Computer Music Conference*, ICMC '07, 2007.

[84] Tom Gross, Chris Stary, and Alex Totter. User-Centered Awareness in Computer-Supported Cooperative Work-Systems: Structured Embedding of Findings from Social Sciences. *International Journal of Human-Computer Interaction*, 18(3):323–360, 2005.

[85] Jonathan Grudin. Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work*, CSCW '88, pages 85–93, New York, NY, USA, 1988. ACM.

[86] Jonathan Grudin. Computer-Supported Cooperative Work: History and Focus. *Computer*, 27(5):19–26, 1994.

[87] Michael Gurevich. JamSpace: A Networked Real-time Collaborative Music Environment. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, pages 821–826, New York, NY, USA, 2006. ACM.

[88] Georg Hadju. Quintet.net: An Interactive Performance Environment for the Internet. In *OSC Conference 2004*, July 2004.

[89] Marc Hassenzahl, Andreas Beu, and Michael Burmester. Engineering Joy. *IEEE Software*, 18(1):70–76, January 2001.

[90] Marc Hassenzahl and Daniel Ullrich. To Do or Not to Do: Differences in User Experience and Retrospective Judgments Depending on the Presence or Absence of Instrumental Goals. *Interacting with Computers*, 19(4):429–437, July 2007.

[91] Mare Hassenzahl, Axel Platz, Michael Burmester, and Katrin Lehner. Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 201–208, New York, NY, USA, 2000. ACM.

[92] Mark Hassenzahl. Attrakdiff: A Service Provided by User Interface Design GMBH. `http://www.attrakdiff.de/en/Services/AttrakDiff-Basic/`, 2007. Accessed November 28, 2012.

[93] Patrick G. T. Healey, Joe Leach, and Nick Bryan-Kinns. Inter-Play: Understanding Group Music Improvisation as a Form of Everyday Interaction. In

Proceedings of Less is More – Simple Computing in an Age of Complexity, 2005.

[94] Mara Helmuth. Virtual Musical Performance and Improvisation on Internet2. *Organised Sound*, 10:201–207, December 2005.

[95] Deborah Hix and H. Rex Hartson. *Developing User Interfaces: Ensuring Usability Through Product and Process*. Wiley professional computing. J. Wiley, 1993.

[96] Kristina Höök, Phoebe Sengers, and Gerd Andersson. Sense and Sensibility: Evaluation and Interactive Art. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 241–248, New York, NY, USA, 2003. ACM.

[97] Wijnand IJsselsteijn, Wouter Van den Hoogen, Christoph Klimmt, Yvonne De Kort, Craig Lindley, Klaus Mathiak, Karolien Poels, Niklas Ravaja, Marko Turpeinen, and Peter Vorderer. Measuring the Experience of Digital Game Enjoyment. In *Proceedings of Measuring Behavior*, 2008.

[98] Katherine Isbister, Kristina Höök, Michael Sharp, and Jarmo Laaksolahti. The Sensual Evaluation Instrument: Developing an Affective Evaluation Tool. In *CHI*, pages 1163–1172, 2006.

[99] Hiroshi Ishii. TeamWorkStation: Towards a Seamless Shared Workspace. In *Proceedings of the 1990 ACM Conference on Computer-supported Cooperative Work*, CSCW '90, pages 13–26, New York, NY, USA, 1990. ACM.

[100] Hiroshi Ishii, Minoru Kobayashi, and Kazuho Arita. Iterative Design of Seamless Collaboration Media. *Communications of the ACM*, 37(8):83–97, August 1994.

[101] Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. Integration of Interpersonal Space and Shared Workspace: ClearBoard Design and Experiments. *ACM Transactions on Information Systems*, 11(4):349–375, October 1993.

[102] Robert Johansen. *GroupWare: Computer Support for Business Teams.* The Free Press, New York, NY, USA, 1988.

[103] R. Burke Johnson and Anthony J. Onwuegbuzie. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, 33(7):14–26, 2004.

[104] Andrew Johnston, Linda Candy, and Ernest Edmonds. Designing and Evaluating Virtual Musical Instruments: Facilitating Conversational User Interaction. *Design Studies*, 29(6):556 – 571, 2008.

[105] Sergi Jordà. FMOL: Toward User-friendly, Sophisticated New Musical Instruments. *Computer Music Journal*, 26:23–39, September 2002.

[106] Ajay Kapur, Ge Wang, Philip Davidson, and Perry R. Cook. Interactive Network Performance: A Dream Worth Dreaming? *Organised Sound*, 10:209–219, December 2005.

[107] Ajay Kapur, Ge Wang, Philip Davidson, Perry R. Cook, Dan Trueman, Tae Hong Park, and Manjul Bhargava. The Gigapop Ritual: A Live Networked Performance Piece for Two Electronic Dholaks, Digital Spoon, DigitalDoo, 6 String Electric Violin, Rbow, Sitar, Tabla and Bass Guitar. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '03, May 2003.

[108] Peter Kauff and Oliver Schreer. An Immersive 3D Video-conferencing System Using Shared Virtual Team User Environments. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments*, CVE '02, pages 105–112, New York, NY, USA, 2002. ACM.

[109] Peter Kauff and Oliver Schreer. Virtual Team User Environments - A Step from Tele-cubicles Towards Distributed Tele-collaboration in Mediated Workspaces. In *International Conference on Multimedia and Expo*, pages 9–12. IEEE, 2002.

[110] Joseph 'Jofish' Kaye, Kirsten Boehner, Jarmo Laaksolahti, and Anna Ståhl. Evaluating Experience-Focused HCI. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2117–2120, New York, NY, USA, 2007. ACM.

[111] Chris Kiefer. A Malleable Interface for Sonic Exploration. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '10, pages 291–296, 2010.

[112] Chris Kiefer, Nick Collins, and Geraldine Fitzpatrick. HCI Methodology For Evaluating Musical Controllers : A Case Study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '08, pages 87–90, 2008.

[113] Kibum Kim, John Bolton, Audrey Girouard, Jeremy Cooperstock, and Roel Vertegaal. TeleHuman: Effects of 3D Perspective on Gaze and Pose Estimation with a Life-size Cylindrical Telepresence Pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2531–2540, New York, NY, USA, 2012. ACM.

[114] David Kirk, Tom Rodden, and Danaë Stanton Fraser. Turn It This Way: Grounding Collaborative Action with Remote Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1039–1048, New York, NY, USA, 2007. ACM.

[115] David Kirsh. Embodied Cognition and the Magical Future of Interaction Design. *ACM Transaction on Computer-Human Interaction*, 20(1):3:1–3:30, April 2013.

[116] Gerhard Kleining. Das Qualitative Experiment. *Kölner Zeitschrift für Soziologie und Sozialpsychologie 38*, 4:724750, 1986.

[117] Dimitri Konstantas, Yann Orlarey, Olivier Carbonel, and Simon Gibbs. The Distributed Musical Rehearsal Environment. *IEEE MultiMedia*, 6(3):54–64, 1999.

[118] Morten Kyng. Designing for Cooperation: Cooperating in Design. *Communications of the ACM*, 34(12):65–73, December 1991.

[119] Craig Latta. Notes from the Netjam Project. *Leonardo Music Journal*, 1(1):103–105, 1991.

[120] Talia Lavie and Noam Tractinsky. Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. *International Journal of Human-Computer Studies*, 60(3):269 – 298, 2004.

[121] Marc Leblanc. 8KindsOfFun.com: The Collected Game Design Rants of Marc "MAHK" LeBlanc. `http://8kindsoffun.com/`. Accessed October 17, 2013.

[122] Jason Lee. The Chopping Board: Real-time Sample Editor. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '06, pages 77–78, 2006.

[123] Min Kyung Lee and Leila Takayama. "Now, I Have a Body": Uses and Social Norms for Mobile Remote Presence in the Workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 33–42, New York, NY, USA, 2011. ACM.

[124] Aaron Levisohn and Thecla Schiphorst. Embodied Engagement: Supporting Movement Awareness in Ubiquitous Computing Systems. *Ubiquitous Learning: An International Journal*, 3(4):97–112, 2011.

[125] Noah Liebman, Michael Nagara, Jacek Spiewla, and Erin Zolkosky. Cuebert: A New Mixing Board Concept for Musical Theatre. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '10, pages 51–56, 2010.

[126] Dan Livingstone and Eduardo Miranda. Composition for Ubiquitous Responsive Sound Environments. In *Proceedings of the International Computer Music Conference*, pages 321–324, 2004.

[127] Lian Loke, Astrid T. Larssen, Toni Robertson, and Jenny Edwards. Understanding Movement for Interaction Design: Frameworks and Approaches. *Personal Ubiquitous Computing*, 11(8):691–701, December 2007.

[128] Lian Loke and Toni Robertson. Inventing and Devising Movement in the Design of Movement-based Interactive Systems. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, OZCHI '08, pages 81–88, New York, NY, USA, 2008. ACM.

[129] Lian Loke and Toni Robertson. Design Representations of Moving Bodies for Interactive, Motion-Sensing Spaces. *International Journal for Human-Computer Studies*, 67(4):394–410, April 2009.

[130] Lian Loke and Toni Robertson. Studies of Dancers: Moving from Experience to Interaction Design. *International Journal of Design*, 4(2):39–54, 2010.

[131] Henrik Lörstad, Mark d'Inverno, and John Eacott. The Intelligent Street: Responsive Sound Environments for Social Interaction. In *Proceedings of the SIGCHI International Conference on Advances in Computer Entertainment Technology*, pages 155–162, New York, NY, USA, 2004.

[132] Matti Luhtala, Tiina Kymäläinen, and Johan Plomp. Designing a Music Performance Space for Persons with Intellectual Learning Disabilities. In Alexander Refsum Jensenius, Anders Tveit, Rolf Inge Godøy, and Dan Overholt, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '11, pages 429–432, Oslo, Norway, 2011.

[133] Craig M. MacDonald and Michael E. Atwood. Changing Perspectives on Evaluation in HCI: Past, Present, and Future. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1969–1978, New York, NY, USA, 2013. ACM.

[134] Raymond MacDonald, Charles Byrne, and Lana Carlton. Creativity and Flow in Musical Composition: An Empirical Investigation. *Psychology of Music*, 34(3):292–306, 2006.

[135] Brian D. Mayton, Nan Zhao, Matt Aldrich, Nicholas Gillian, and Joseph A. Paradiso. WristQue: A personal sensor wristband. In *Body Sensor Networks (BSN), 2013 IEEE International Conference on*, pages 1–6, May 2013.

[136] Martin Miller. KVR: VSTunnel by Martin Miller. `http://www.kvraudio.com/product/vstunnel_by_martin_miller`. Accessed October 28, 2013.

[137] Kevin L. Mills. Computer-Supported Cooperative Work Challenges. In *Encyclopedia of Library and Information Sciences*, pages 666–677. Marcel Dekker, 2003.

[138] Michael J. Muller. Participatory Design: The Third Space in HCI. In Julie A. Jacko and Andrew Sears, editors, *The Human-Computer Interaction Handbook*, pages 1051–1068. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003.

[139] Thomas Nagel. *The View From Nowhere*. Oxford University Press, USA, 1986.

[140] Thomas Nagel. Against the Quantitative-Qualitative Incompatability Thesis, or, Dogmas die hard. *Educational Researcher*, 17, 1988.

[141] Jakob Nielsen. Risks of Quantitative Studies. `http://www.useit.com/alertbox/20040301.html`, March 1 2004. Accessed November 16, 2012.

[142] Donald A. Norman. *The Design of Everyday Things*. Basic Books, New York, reprint paperback edition, 2002.

[143] Donald A. Norman and Stephen W. Draper. *User Centered System Design: New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1986.

[144] Heather L. O'Brien and Elaine G. Toms. The Development and Evaluation of a Survey to Measure User Engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69, January 2010.

[145] Nicola Orio, Norbert Schnell, and Marcelo M. Wanderley. Input devices for musical expression: borrowing tools from HCI. In *Proceedings of International*

Conference on New Interfaces for Musical Expression, NIME '01, pages 1–4, Singapore, Singapore, 2001. National University of Singapore.

[146] François Pachet, Oliver Delerue, and Peter Hanappe. Dynamic Audio Mixing. In *Proceedings of International Computer Music Conference*, ICMC '00, Berlin, 2000. ICMA.

[147] François Pachet and Olivier Delerue. MusicSpace: a Constraint-based Control System for Music Spatialization. In *Proceedings of ICMC 1999*, pages 272–275, Beijing, 1999. ICMA.

[148] Victor M. R. Penichet, Maria D. Lozano, José A. Gallud, and Ricardo Tesoriero. User Interface Analysis for Groupware Applications in the TOUCHE Process Model. *Advances in Engineering Software*, 40(12):1212–1222, December 2009.

[149] Marianne Graves Petersen, Ole Sejer Iversen, Peter Gall Krogh, and Martin Ludvigsen. Aesthetic Interaction: A Pragmatist's Aesthetics of Interactive Systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, DIS '04, pages 269–276, New York, NY, USA, 2004. ACM.

[150] Cornelius Poepel. On Interface Expressivity: A Player-Based Study. In *Proceedings of the International Conference on New interfaces for Musical Expression*, NIME '05, pages 228–231, Singapore, Singapore, 2005. National University of Singapore.

[151] F. Poletti, N. V. Wheeler, M. N. Petrovich, N. Baddela, E. Numkam Fokoua, J. R. Hayes, D. R. Gray, Z. Li, R. Slavík, and D. J. Richardson. Towards high-capacity fibre-optic communications at the speed of light in vacuum, 2013.

[152] Amandine Pras and Catherine Guastavino. Improving the Sound Quality of Recordings Through Communication Between Musicians and Sound Engineers. In *Proceedings of the International Computer Music Conference*, August 2009.

[153] Amandine Pras and Catherine Guastavino. The Impact of Producers' Comments and Musicians' Self-Evaluation on Performance During Recording Sessions. In *Audio Engineering Society Convention 131*, 10 2011.

[154] Amandine Pras and Catherine Guastavino. The Role of Music Producers and Sound Engineers in the Current Recording Context, as Perceived by Young Professionals. *Musicae Scientiae*, 15(1):73–95, 2011.

[155] Pamela Ravasio, Sissel Guttormsen-Schar, and Vincent Tscherte. The Qualitative Experiment in HCI: Definition, Occurrences, Value and Use. *Transactions on Computer-Human Interaction*, pages 1–24, 2004.

[156] Pedro Rebelo, Perdro Carneiro, and Franziska Schroeder. A-Synk. `http://www.somasa.qub.ac.uk/~prebelo/index/works/asynk/left.htm`. Accessed May 15, 2011.

[157] Pedro Rebelo and Alain B. Renaud. The Frequencyliator: Distributing Structures for Networked Laptop Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '06, pages 53–56, Paris, France, France, 2006. IRCAM Centre Pompidou.

[158] Alain Renaud. Dynamic Cues for Network Music Interaction. *7th Sound Music and Computing Conference*, July 2010.

[159] Alain B. Renaud and Pedro Rebelo. Network Performance: Strategies and Applications. *Proceedings of the Conference on New Interfaces for Musical Expression*, 2006.

[160] Charlie Roberts. Multi-touch, Consumers and Developers. Technical report, Media Arts and Technology Program, University of California, 2008.

[161] Tom Rodden and Gordon Blair. CSCW and Distributed Systems: The Problem of Control. In Liam Bannon, Mike Robinson, and Kjeld Schmidt, editors, *Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW '91*, pages 49–64. Springer Netherlands, 1991.

[162] Gerard Roma and Anna Xambó. A Tabletop Waveform Editor for Live Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, NIME '08, pages 249–252, 2008.

[163] Manfred Rürup. Digital Music Link. `http://www.digitalmusician.net`. Accessed July 21, 2011.

[164] Christopher Salter, Marije Baalman, and Daniel Moody-Grigsby. Between Mapping, Sonification and Composition: Responsive Audio Environments in Live Performance. In *Computer Music Modeling and Retrieval: Sense of Sounds*, pages 246–262. Springer-Verlag, Berlin, Heidelberg, 2008.

[165] Allison Sauppé and Bilge Mutlu. How Social Cues Shape Task Coordination and Communication. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 97–108, New York, NY, USA, 2014. ACM.

[166] Alexander Sawchuk, Elaine Chew, Roger Zimmermann, Chris Papadopoulos, and Chris Kyriakakis. From Remote Media Immersion to Distributed Immersive Performance. In *Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence*, ETP '03, pages 110–120, New York, NY, USA, 2003. ACM.

[167] Fransizka Schroeder, Alain B. Renaud, Pedro Rebelo, and Fernando Gualda. Addressing the Network : Performative Strategies for Playing Apart. In *International Computer Music Conference*, 2007.

[168] Franziska Schroeder and Rebelo Pedro. Sounding the Network: The Body as Disturbant. *Leonardo Electronic Almanac*, 16, 2009.

[169] Nate Schuett. *Interconnected Musical Networks: Bringing Expression and Thoughtfulness to Collaborative Music Making*. PhD thesis, Stanford Center for Computer Research in Music and Acoustics, 2002.

[170] Abigail Sellen, Bill Buxton, and John Arnott. Using Spatial Cues to Improve Videoconferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 651–652, New York, NY, USA, 1992. ACM.

[171] David Sirkin and Wendy Ju. Consistency in Physical and On-screen Action Improves Perceptions of Telepresence Robots. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pages 57–64, New York, NY, USA, 2012. ACM.

[172] Carolyn Snyder. Paper Prototyping. `http://www.snyderconsulting.net/article_paperprototyping.htm`. Accessed August 22, 2011.

[173] Vibeke Sorensen. Global Visual Jam Project. In *Leonardo*, volume 38, page 316. MIT Press, 2005.

[174] Dominik Strohmeier, Satu Jumisko-Pyykkö, and Kristina Kunze. Open Profiling of Quality: A Mixed Method Approach to Understanding Multimodal Quality Perception. *Advances in Multimedia*, 2010, 2010.

[175] Nicolas Sturmel, Antoine Liutkus, Jonathan Pinel, Laurent Girin, Sylvain Marchand, Gal Richard, Roland Badeau, and Laurent Daudet. Linear Mixing Models for Active Listening of Music Productions in Realistic Studio Conditions. In *Audio Engineering Society Convention 132*, Apr 2012.

[176] Penelope Sweetser and Peta Wyeth. GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.*, 3(3):3–3, July 2005.

[177] Koray Tahiroglu. Dynamic Social Interaction in a Collective Mobile Music Performance. In *Proceedings of the 2009 International Conference on Computational Science and Engineering*, pages 854–859, Washington, DC, USA, 2009.

[178] Atau Tanaka. Network Audio Performance and Installation. In *Proceedings of International Computer Music Conference*, ICMC '99, 1999.

[179] Atau Tanaka. Musical Performance Practice on Sensor-based Instruments. In Marcelo M. Wanderley and Marc Battier, editors, *Trends in Gestural Control of Music*, pages 389–405. IRCAM Centre Pompidou, 2000.

[180] Atau Tanaka. Musical implications of Network Infrastructures: Perturbations of traditional artistic roles. In Jean-Piere Balpe and Sylvie Leleu-Merviel, editors, *Actes de H2PTM01, Hypertextes/Hypermedias: Nouvelles ecritures, nouveaux langages*, pages 241–250. Hèrmes Science Publications, 2001.

[181] Atau Tanaka. Interaction, Experience and the Future of Music. In Dan Diaper, Colston Sanger, Kenton O'Hara, and Barry Brown, editors, *Consuming Music Together*, volume 35, pages 267–288. Springer Netherlands, 2006.

[182] Atau Tanaka and Bert Bongers. Global String: A Musical Instrument for Hybrid Space. In *Proceedings of Cast01: Living in Mixed Realities*, 2001.

[183] John C. Tang, Robert Xiao, Aaron Hoff, Gina Venolia, Patrick Therien, and Asta Roseway. HomeProxy: Exploring a Physical Proxy for Video Communication in the Home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1339–1342, New York, NY, USA, 2013. ACM.

[184] Michael Terry and Elizabeth D. Mynatt. Recognizing Creative Needs in User Interface Design. In *Proceedings of the 4th Conference on Creativity & Cognition*, C&C '02, pages 38–44, New York, NY, USA, 2002. ACM.

[185] Gary Thompson. Personas. `http://wiki.fluidproject.org/display/fluid/Personas`. Accessed August 24, 2011.

[186] Noam Tractinsky, A. S. Katz, and D. Ikar. What is Beautiful is Usable. *Interacting with Computers*, 13(2):127–145, 2000.

[187] Marc Vass, John M. Carroll, and Clifford A. Shaffer. Supporting Creativity in Problem Solving Environments. In *Proceedings of the 4th Conference on*

Creativity & Cognition, C&C '02, pages 31–37, New York, NY, USA, 2002. ACM.

[188] Karel Vredenburg, Ji-Ye Mao, Paul W. Smith, and Tom Carey. A Survey of User-Centered Design Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 471–478, New York, NY, USA, 2002. ACM.

[189] Marcelo Mortensen Wanderley and Nicola Orio. Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Comput. Music J.*, 26(3):62–76, September 2002.

[190] Gil Weinberg. The Aesthetics, History, and Future Challenges of Interconnected Music Networks. *Proceedings of the 2002 International Computer Music Conference*, pages 349–356, 2002.

[191] Gil Weinberg. Interconnected Musical Networks: Toward a Theoretical Framework. *Computer Music Journal*, 29:23–39, June 2005.

[192] Charlotte Wiberg. Affective Computing vs. Usability? Insights of Using Traditional Usability Evaluation Methods. In *CHI 2005 Workshop on Innovative Approaches to Evaluating Affective Interfaces*, April 2005.

[193] William J. Wrigley and Stephen B. Emmerson. The Experience of the Flow State in Live Music Performance. *Psychology of Music*, 2011.

[194] Seionshin Yamagishi. Variations for WWW: Network Music by MAX and the WWW. In *Proceedings of the 1998 International Computer Music Conference*, pages 510–513, 1998.

[195] Indi Young. *Mental Models: Aligning Design Strategy with Human Behavior*. Rosenfeld Media, Brooklyn, New York, 1st edition, 2008.

[196] John Yu. Computer Generated Music Composition. Master's thesis, Massachusetts Institute of Technology, 1996.

Appendix A

Full Report by Steve Cowan, Composer

My time with the performance system currently in development by Dalia El-Shimy was an eye-opening and fun experience. With various music-making programs and devices becoming more affordable and readily available to all musicians, as well as modern developments in motion sensory technology with respect to things such as videogame systems, the idea of combining these two worlds is one that is full of potential.

My original understanding of the system was that it was meant to connect multiple musicians playing together, all in different locations or rooms. Each player would then have the ability to manipulate the mixes or levels of what they are hearing, through motion detection software and other various features. As a music teacher myself who has experimented with teaching via Skype, this idea has obvious benefits of connecting people who can't otherwise be in the same physical space, but also comes with the inherent issue of lag and stable Internet connection.

So, I entered the experiment as a solo musician with the hopes of refining the features and finding other useful situations for the system. I did this with the aid of another piece of technology that I use quite often, called a "Loop Station". This device allows me to have pre-recorded tracks playing while I perform on an electric guitar in real time. With separate outputs, I then plugged the different tracks into the system individually, thus creating an environment that simulated a 3-person interactive experience. This allowed me to experiment with the features more efficiently, as I would not have to coordinate little experiments with other people and could control it all from the interface in front of me. It also gave me a new perspective of how the system could be used by an individual, or home studio "producer", a music making approach that is more popular now than ever.

This approach led to certain realizations about how the system can be used, and what situations it would be most useful in. While it could work in a live in-studio recording session, or perhaps rehearsing material that the musicians are already familiar with, the motion-manipulation seems to offer the most advantages to musicians who are in the early stages of the creative process.

For example, if a small group of musicians have a rough sketch for a new piece of music planned out but the specifics of individual parts need refining, the means of manipulating what one hears through simple body motions is extremely convenient. If player 1 is trying out some ideas and wants to hear more or less of the other players, they can simply walk forward or backward instead of stopping the music and manually adjusting some faders. This allows for each musician to isolate either their part, or the other musician's parts, through simple body motions. They will hear how all different mixing approaches or different volume levels can affect the final product - and thus will influence the creative decisions they make in the process.

As a single musician, these features basically helped cut out a tedious process that I often avoid. What I mean is this: almost every musician I know these days has some sort of recording software on their computer, and thus has the ability to record and produce multi-track recordings at home. Personally, I find all the clicking and computer-based activity in this to drain my creative energy and make the process frustrating. As soon as I got my previously mentioned Loop Station, and thus was able to record multiple tracks instantly by simply tapping my foot on this pedal, my creative output skyrocketed.

On that note, even though I was creating much more original music and at a higher caliber, I was still avoiding the next step of recording, mixing, and making the final product more presentable to a typical listener. This is because the click-based activities of controlling things like levels / panning within recording software was equally unappealing to me.

Using the performance system here, I was able to get some great solutions for these issues without having to do anything other than play my music in real time, and move my body a bit. I was easily able to see which tracks sounded best panned left, or right, or in the center; I was able to hear which textures were better off in the foreground, and which sounded better off more "distant", perhaps with a hint of reverb; I was able to iron out how two musical ideas interacted one on one, and then with a slight 90 degree turn, could hear how it then sounded with a third musical idea in the mix. It provided more clarity and depth than what my Loop Station offers as far as coming up with new ideas, and also, gave me insight on how to start the "mixing" or post-creative production aspect of releasing recordings. With the system working hand-in-hand with a recording program, some people with undoubtedly take it a step further and have the "mix" and "levels" be determined within a live-performance, based on how the performer moved during the recording.

Other than dynamic manipulations to volume and reverb, the three features I worked with also provided a logical succession for the creative process. Track panning allows the ability to work on ideas one on one, by cutting out one of the 3 musicians with a simple torso pivot. The mix control brings all 3 players into the mix, but with the ability to pan your own part around to see how everything is blending/working together. Then the spatialization is a good final step, fleshing out the music ideas into their own space within the panning, and hearing how it works in a situation that will sound closer to the eventual desired final product (be it a live performance or a recording), while also having some manipulations available to you if you walk around within the simulated space.

In conclusion, the features that this system offered were fun, useful, and helped me come up with new musical and production ideas. However, for use by a singular musician, it needed to be used in combination with something else. I happened to have my Loop Station, but for those who don't have such a pedal, the system would have to be some sort of add-on available to various recording software. Another obvious criticism would be its use for musicians or instrumentalists who are required to be more stationary while playing. Luckily, electric guitar is easy and natural to move around. Other motion-detecting approaches would have to be put in place for a keyboardist, for example.