Designing Experiential Media for Volitional Usage:

An Approach Based on Music and Other Hobbies

by

Isaac Wallis

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved April 2013 by the
Graduate Supervisory Committee:

Todd Ingalls, Chair
Grisha Coleman
Hari Sundaram

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

Achievement of many long-term goals requires sustained practice over long durations. Examples include goals related to areas of high personal and societal benefit, such as physical fitness, which requires a practice of frequent exercise; self-education, which requires a practice of frequent study; or personal productivity, which requires a practice of performing work. Maintaining these practices can be difficult, because even though obvious benefits come with achieving these goals, an individual's willpower may not always be sufficient to sustain the required effort. This dissertation advocates addressing this problem by designing novel interfaces that provide people with new practices that are fun and enjoyable, thereby reducing the need for users to draw upon willpower when pursuing these long-term goals. To draw volitional usage, these practice-oriented interfaces can integrate key characteristics of existing activities, such as music-making and other hobbies, that are already known to draw voluntary participation over long durations.

This dissertation makes several key contributions to provide designers with the necessary tools to create practice-oriented interfaces. First, it consolidates and synthesizes key ideas from fields such as activity theory, self-determination theory, HCI design, and serious leisure. It also provides a new conceptual framework consisting of heuristics for designing systems that draw new users, plus heuristics for making systems that will continue drawing usage from existing users over time. These heuristics serve as a collection of useful ideas to consider when analyzing or designing systems, and this dissertation postulates that if designers build these characteristics into their products, the resulting systems will draw more volitional usage. To demonstrate the framework's usefulness as an analytical tool, it is applied as a set of analytical lenses upon three previously-existing experiential media systems. To demonstrate its usefulness as a design tool, the framework is used as a guide in the development of an experiential media system called *pdMusic*. This system is installed at public events for user studies, and the study results provide qualitative support for many framework heuristics. Lastly, this dissertation makes recommen-

dations to scholars and designers on potential future ways to examine the topic of volitional

usage.

For family.

ACKNOWLEDGEMENTS

I could not have finished this dissertation without the help of Prof. Todd Ingalls, Dr. Hari Sundaram, Prof. Grisha Coleman, Dr. Ellen Campana, Andrea Silkey, Dr. Lisa Tolentino, Ryan Spicer, Dr. Alex Fink, Diana Siwiak, Nicole Williams, Kristi Rogers, Prof. Mary Neubauer, Cathy Vuong, Janel Goodman, Nicci Lehrer, Mike Baran, Dr. Daragh Byrne, Dr. Thanassis Rikakis, Dr. Garth Paine, Byron Lahey, Dr. Margaret Duff, Loren Olsen, and many others. Thank you all for your time, effort, thought, and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Twelve years ago, I suffered an injury to my spinal cord in an automobile accident. The disc between my fifth and sixth cervical vertebrae herniated, bulging against my spinal cord. The resulting spinal contusion caused temporary paralysis in all four limbs for a period of approximately one month. Then, after surgery, my strength returned gradually over the course of a year. My right arm was the slowest to rehabilitate; long after I was walking again, I still could not lift my arm above my waist or wiggle my fingers. However, as my spinal cord slowly healed, these symptoms too began to decrease. I was an amateur musician prior to the accident, and I knew music-making would be the activity I missed most if I never regained use of my arm. As soon as I was capable of pinching my thumb and first finger together strongly enough to hold a pick, I immediately set about relearning the guitar, an instrument with which I had gained some proficiency in the years prior to the accident. I practiced daily. A few months later, my right hand had regained functionality to the point where I could return to my childhood instrument, the piano, which I also practiced daily. Eventually, I achieved proficiency such that I could enroll in music college, and there I practiced instruments even more.

I relate this story as an illustration of the power of hobby. Ten years after my injury, there are only trace reminders of nerve damage. There is no way to know how much of my successful rehabilitation is due to my musical hobby and how much would have happened without practicing instruments. At the time, however, my doctors and rehabilitation specialists told me that my instrumental practice could only be beneficial to recovery, and I did have an unusually successful rehabilitation from this sort of spinal cord injury. Many people with these injuries never regain the full use of their limbs. Without my hobby, the recovery may not have been so successful. Because of this experience, at some point I began to wonder: were there any

1

other ways in which hobbies and passions of mine had helped me in the course of my life? For example, were there any music theory concepts I learned, for fun, that may have prepared me for mathematical concepts that were taught to me in school? What about non-musical hobbies; how had they helped me? I became very interested in the topic of hobbies: why they are taken up, what makes them so addictive, and how they benefit people.

From the literature it seems clear that hobbies do benefit people. For example, Gee (2007, ch. 9) discusses reading levels amongst grade school children who play and enjoy *Yu-Gi-Oh*[1], a trading card game. *Yu-Gi-Oh* contains a pool of over 10,000 available cards, and each card is different. Upon each card is printed a short summary of the actions and counter-actions afforded by the card; effectively, each card contains its own rules of play. Players acquire cards from this pool to build highly personalized decks of 40–60 cards. Different deck compositions afford different strategies of play, so there is incredibly high strategic complexity involved with regard to both game-play and deck-building. What is interesting is that, as Gee notes, although the recommended age for *Yu-Gi-Oh* players is 6 and up, the text on the cards is "in complexity, far above the language many young children see in their schoolbooks, until they get to middle school at best and, perhaps, even high school. But seven-year-old children deal and deal well with this language" (pg. 111). Playing *Yu-Gi-Oh* causes children to practice and master advanced reading and language skills at early ages. Similar educational effects can be seen with the hobby of playing with LEGO bricks. For twenty-five years, MIT professor Turkle assigned a short essay as the first assignment in her classes on the social studies of science and technology. Each student was asked to write about an object from his or her childhood that impacted how the student came to be in Turkle's class. In Turkle (2008), she compiles 51 of these essays, and includes eight similar essays written by well-known and respected senior scientists, engineers, designers, and technologists. LEGOs were brought up in so many of these essays that Turkle devotes a section to them, claiming: "Over the years, so many students have chosen LEGOs as

---

[1] see www.yugioh-card.com/en/

the key object on their path to science that I am able to take them as a constant to demonstrate the wide range of thinking and learning styles that constitute a scientific mindset" (pg. 7).

In an ideal world, when I was healing from my car accident I might have found my therapist-prescribed rehabilitation exercises as fun and engaging as I did my guitar. Unfortunately, I did not; I frequently ignored my physical therapist's exercise recommendations. Now, if you were to extend my experience across the planet, you would probably find many other people who similarly find rehabilitative exercises tedious and skip them, but do not play music. This contributes to the economic cost of non-compliance to medical interventions; which is, according to a review in the Journal of Clinical Pharmacy and Therapeutics, on the order of 100 billion dollars per year in the Unites States (Vermeire et al., 2001). If, hypothetically, we could create interactive media systems—systems that respond to user actions by presenting digital media content such as audio or graphics—that helped patients perform rehabilitative exercises, and were fun enough that patients would use them voluntarily in their free time, we would be addressing a huge societal problem. By extension, we would be reducing a huge economic drain on healthcare systems world-wide. Benefits at a similar societal scale would be accrued if practice-oriented interfaces were designed for educational applications, physical fitness applications, productivity applications, or other applications that require individuals to make effort over time.

Obviously, people are capable of doing exercise, staying productive, educating themselves, and rehabilitating from various maladies without the help of any interactive media at all. However, people often fail to do so. It seems that although they recognize the benefits of accomplishing their goals, and they wish to receive those benefits, they find it taxing to sustain the long-term practices that are required to accomplish these goals. Therefore, reducing the tedium of performing these practices is where interactive media could be of great benefit. Products are beginning to appear in the marketplace that are based on this idea. For example, the Wii

Fit[2] suite of games is designed to simultaneously: a) be fun, engaging, and marketable; and b) help people reach their fitness goals. At present, the topic of serious gaming is burgeoning in academia, and companies such as Sifteo[3] and Leapfrog[4] are also capitalizing on this concept by producing and marketing educational video games. Interactive media for language learning, such as Rosetta Stone[5] also works toward this end. Interactive systems are beginning to be deployed in rehabilitative therapy. One example of this is work by Rizzo et al. (2005), describing a simulation used to treat PTSD in soldiers; another example is the stroke rehabilitative system described in chapter 4 of this dissertation. Many other examples exist of systems designed to help users sustain practices in the effort of achieving various goals. In this dissertation, these interactive systems are called 'practice-oriented interfaces.' Practice-oriented interfaces help people make beneficial self-transformations.

In order for practice-oriented systems to succeed, they must be used and used repeatedly. Unfortunately, designers seem to find it difficult to make systems that inspire this sort of repeated and engaged usage. Users seem to find many new interactive interfaces novel in the short run but boring or frustrating over time, and a corresponding drop-off in usage is the result. There is no concise set of generalizable concepts or rules to help guide designers in the making of systems that depend on long-term volitional usage—extensive searches for such a framework in libraries like the Association for Computing Machinery (ACM) library yielded no results. Video games have had some success in drawing long-term volitional usage, and this might explain why most of the products listed in the previous paragraph are game-like. But research on volitional usage should not be entirely centered on serious gaming and gamification, because a) many people do not find video games compelling, and b) games cannot be designed to appropriately address the needs of all users with regard to their goals. More to the point,

---

[2] see wiifit.com

[3] see www.sifteo.com

[4] see www.leapfrog.com

[5] see www.rosettastone.com

games are only one activity among a huge number of involved activities that people participate in, over long time periods, for the purposes of leisure. Here I name this wider set of activities 'hobby.' Hobbies are important to hobbyists and are beneficial to their lives in many ways, in no small part through adding enjoyment and meaning.

Many existing hobbies require no interactive media, but in the future, I believe that many new hobbies will be based in digital media and new interfaces. I contend that a need exists for designers of these new hobbies. In other words, I believe that designers should know how to make interactive media that can be enjoyed and adopted in hobby-like ways. Such interaction designers have the opportunity to create practice-oriented interfaces that provide new hobbies while simultaneously motivating socially beneficial practices. Therefore, the fact that the field of interaction design has not mined the subject of hobby for ideas and theory in any systematic way—currently, searching the ACM library for publication abstracts containing the terms "hobby" or "serious leisure" yields only 25 results (to illustrate how minuscule this amount is, searching for "gaming" yields 385,120 results)—represents a significant oversight of the field. This dissertation exists, in part, to begin addressing this oversight.

In this dissertation I work toward improving the ability for interaction designers to make practice-oriented interfaces that draw volitional usage over long time periods. I believe hobbies that draw participation do so because they are well-designed to draw participation; they share characteristics that people find appealing. Therefore, this dissertation contributes a design framework—a set of concepts for designers to use as guidelines when analyzing or designing practice-oriented systems for volitional usage—built around these appealing hobby character-istics. The concepts in the framework can be used as a set of analytical lenses to identify areas of potential improvement with regard to volitional usage in existing systems, and they can also be used as rules-of-thumb when designing entirely new systems. In addition to this key contribution, the following contributions are also made:

- A consolidated exploration of literature relevant to the topic of volitional usage from many diverse fields of study including arts and aesthetics, human-computer interface design, and the psychology of motivation and emotion. Key ideas from this literature form the bases for many aspects of the design framework.

- Three discussions about earlier interactive media systems I helped design, each offering valuable insights into different facets of volitional usage. The experience of working on these systems informed certain aspects of the design framework. In order to demonstrate the framework's usefulness as an analytical tool, these discussions include framework-based analyses.

- Discussion about a new practice-oriented system that was designed using the framework. Various studies were performed on the users of this system during various stages of its development. These studies were used to inform and fine-tune aspects of the framework, and they also provided data supporting aspects of the framework.

It is important, at the outset of any large-scale work of scholarship such as this dissertation, to outline potential biases in an upfront way. This dissertation was written primarily in the context of multimodal interface design, which is the type of design in which I have specialized during the course of my studies. Specifically, I have specialized in the design of multimodal interfaces containing: a) sensing that includes computerized tracking of bodily motions, and b) interactive feedback that includes auditory elements such as sonification or music. The interactivity is driven by human movement, and the *raison d'être* for the systems is the movement itself. Some of the systems I have worked on were designed for rehabilitative purposes, and others were designed for educational or artistic purposes. Everything in this dissertation stems from my work on these types of systems. It is likely that the examples I use, the systems I analyze and design, and the framework I give would look differently if addressed by someone coming from a background in robotics, visual arts, psychology, or any of a number of other

fields. However, it is my hope and intention that designers from these fields will be able to find value pertaining to their own work here.

During the course of my studies, I have published a number of papers, in conference proceedings and as book chapters, on various topics related to this dissertation (cf. Wallis et al., 2007, 2008, 2011a,b, 2013). One of the personal benefits of this dissertation, for me, is that it synthesizes these previous publications into a single theoretical perspective and narrative. Permission has been granted from the copyright holders—in most cases, I retained the copyright—to incorporate these publications into this work. I endeavor to cite areas where these papers are used, but they have been broken up, reformulated, and rewritten for the purposes of dissertation cohesion and language flow—to the extent that now, pieces of them are scattered throughout, and it would be impractical to cite the source wherever one of these snippets of writing are used. In sections that depend heavily on language from one of these publications, I cite the relevant publication near the top of the section.

Here in chapter 1, I have discussed the relevance of this dissertation topic, explaining why work in the field of volitional usage is beneficial to society. The rest of this dissertation is organized as follows: In chapter 2, "Approaches," I provide a literature review, define terminology, and synthesize research from various scholarly domains that is relevant to the topic of volitional usage. In chapter 3, "The Framework," I give the framework which is the ultimate goal of this dissertation. In chapter chapter 4, "The Framework as Analytical Tool," I discuss previous interactive systems that were designed in the course of my studies, and how they relate to the topic of volitional usage. Because these systems preceded the framework, they can not be formally evaluated with regard to the framework; however, working on them helped me formulate the framework, and the framework can be applied on them analytically for the purpose of understanding volitional usage in their specific circumstances. In chapter 5, "The Framework as Design Tool," I document and describe a new movement-based system developed for volitional usage which is created using the framework, and also discuss aspects

7

of its formal evaluation. In chapter 6, "Future Work," I discuss how the science of volitional usage can continue to be furthered, including brief discussion of upcoming practice-oriented systems that are in different stages of development. Lastly, in chapter 7, "Conclusion," I discuss some of the challenges and deficiencies of the work given here, and discuss its implications for society and the future of interactive systems design.

Chapter 2

APPROACHES

An interdisciplinary approach is required to address the topic of volitional usage in interactive media. Literature exists within various academic domains addressing aspects of the topic, but the ideas from these domains have yet to be synthesized into a cohesive theory that can be used as a reference for designers of interactive systems. In this chapter, I explore ideas from the literature of three of the most relevant academic domains—the arts, design, and psychology. This exploration results in discussions on the artistic and aesthetic nature of activity and hobby (section 2.1), the design of human-computer interfaces for the creation of experiences (section 2.2), and theories of emotion, engagement, and motivation (section 2.3). These discussions define terminology and explain various approaches underlying my work. They will facilitate understanding in other aspects of this dissertation, such as the design framework presented in chapter 3 or elements of system designs in chapter 4.

2.1 Activity, Hobby, and Aesthetics

This section exists to outline foundational interaction design philosophies that underpin numerous decisions in this dissertation. Starting with a wide scope, I argue that interactive systems exist to afford or facilitate activities. Given this basis, it stands to reason that we can look at existing (non-mediated) activities for insight into interaction design. Hobbies are activities that are exemplary with regard to volitional participation, so understanding hobbies is important when designing for volitional usage—one hobby that is frequently used as a template for interactive systems design in this dissertation is music-making. Designing things for the purpose of affording new activities is a creative act, and like other creative acts such as art-making, can be discussed in aesthetic terms. In this section I discuss how the language of aesthetics might apply to the design of practice-oriented interfaces, and argue that the design of practice-oriented interfaces is akin to a new art-form.

Activity theory holds that the structure and content of our minds are intrinsically entangled with the activities we perform (Kaptelinin and Nardi, 2012). We do activities in order to meet our needs, but it turns out that there is a reciprocal relationship: we internalize something in the process, so the activity also operates on us. According to the theory's originator (cf. Leontyev, 1981), the performance of activities is responsible for the production of the human mind, both in the historical sense (i.e. moving from primordial minds to modern minds), and in the individual sense (i.e. moving from infant mind to adult mind). If action affects the mind, and the mind controls action, human activity becomes a medium for growth. By doing things, we change who we are, which in turn changes the sort of things we will do in the future. This dissertation is concerned with the question of how to design systems that encourage users to invest in doing things.

Therefore, I take the approach advocated by Norman (2005), and treat interactive media as nothing more than vehicle for activity. This treatment has design implications. For

example, if a designer is making an interactive system and wants it to be engaging, then one of the first steps in his or her design process must be to choose an engaging activity that the interactive system will afford. This will determine the entire form of the system. People do not become engaged with interactive systems, they become engaged with activities those systems provide. This is true no matter what sort of interactive technology is being discussed. For example, consider the question: Is television engaging? The screen itself is not—no one would voluntarily stare at a blank screen, for example—but the activity of watching theatrical narratives is. Television merely affords the activity of watching theatrical narratives. In this dissertation, I will make frequent reference to the 'afforded activities' that interfaces provide.

2.1.1   Hobby

Activity theory explains, in part, the motivating power of hobby: when we take part in an activity, we invest in it, and the more we invest, the more it becomes part of who we are. Hobbies are central to this dissertation because they are among the best examples of volitional participation in day-to-day life. If one excludes passive forms of entertainment such as television watching, hobbies are *the* best example, because people regularly participate in their hobbies over the course of years with no incentive other than fun and relaxation. They are enthusiastic enough to participate during leisure time without regard for profit (although there may be aspirations of eventually making the hobby into a career). Whether the hobbies are based in pursuits such as woodworking, poetry-writing, stamp-collecting, hockey-playing, or playing online video games, improving skill-level is enjoyable in itself to hobbyists. In his seminal treatise on aesthetics, "Art as Experience" (1934), Dewey wished that a word existed combining 'artistry' (producing art) with 'aesthetics' (appreciating art). He considered these elements inseparable, since great artists create works for their audience's appreciation, and audiences need an understanding of the intricacies of production to fully appreciate the works. Such a term would also be useful in discussions of hobbyists, because hobbyists often exemplify

the combination between production and appreciation of aesthetic work. They do appreciate the outcomes of their work, and try to ensure those outcomes are aesthetically pleasing when applicable, and enjoy sharing those outcomes with other people—but in many cases, the only audiences are the hobbyists themselves. Therefore, the real impetus behind a hobby practice is the act of practicing itself, and this explains why many hobbyists seem satisfied to pursue their hobbies in solitude. For example, many amateur poets write pieces only for themselves, with no intention of sharing their work.

Hobbies are activities that draw regular volitional usage over long time periods. Myriad activities exist in the world that fit this definition, and they vary so dramatically in form that even just classifying the various hobbies is a complex and challenging task. The best attempt for this comes from a field of sociological scholarship called leisure studies. Figure 2.1 shows a well-known categorization of leisure from that field. In this categorization, casual leisure includes the passive forms of entertainment such as watching TV, project-based leisure includes one-off or infrequent projects such as making a birthday gift, and devotee work consists of paid work in fields populated by passionate professionals. The category of 'serious leisure' and its sub-categories are most relevant to this dissertation.

I use the term 'hobby' as a convenient short-hand for what Stebbins defines as 'serious leisure.' I contend that the terms are equivalent: even though Stebbins considers hobbyism a subcategory that is separate from amateurism and volunteerism, it is accurate to call amateurs and volunteers hobbyists. For example, amateur musicians can be said to have a hobby of playing music, and volunteers can be said to have a hobby of helping people. Therefore, the definition for hobby, using the definition for serious leisure in Stebbins (1982, pg. 5), is "the systematic pursuit of an amateur, hobbyist, or volunteer core activity that people find so substantial, interesting, and fulfilling that, in the typical case, they launch themselves on a leisure career centered on acquiring a combination of its special skills, knowledge, and experience." The term 'leisure career' denotes the life-cycle of a person's involvement with a specific serious

The Serious Leisure Perspective
(version February 2013)

Casual Leisure
- play
- relaxation
- passive entertainment
- active entertainment
- sociable conversation
- sensory stimulation
- casual volunteering
- pleasurable aerobic activity

Project-Based Leisure
- One-shot Projects
  - making/tinkering
  - liberal arts
  - activity participation
  - volunteering
  - arts projects
- Occasional Projects

Leisure

Serious Pursuits

Serious Leisure
- **Amateur**
  - art
  - science
  - sport
  - entertainment
- **Volunteer**
  - popular
  - idea-based
  - material
  - floral
  - faunal
  - environmental
- **Hobbyist**
  - collecting
  - making & tinkering
  - activity participation
  - sports & games
  - liberal arts pursuits
- **Devotee Work**
  - liberal professions
  - consulting/counselling occupations
  - some skilled trades
  - some small businesses

Diagram formulated by Jenna Hartel

Figure 2.1: Stebbins' categorization of serious leisure. From the Serious Leisure Perspective Website, www.seriousleisure.net. (February, 2013 version)

leisure activity. For example, one might say a beginning guitarist is early in his or her leisure career, or that a former stamp collector has ended a leisure career. Since the time-frames of involvement with serious leisure are comparable to that of professional careers, this seems a useful formulation. Ergo, when I discuss the life-cycle of system usage in this dissertation, I will sometimes use the phrase 'usage career.'

Thus, according to the serious leisure categorization shown in fig. 2.1, the activities that are most relevant to this dissertation are: amateur art, science, sport, or entertainment; popular, idea-based, material, floral, faunal, or environmental volunteerism; or hobbies based in collecting, making and tinkering, activity participation, sports and games, or liberal arts pursuits. It is important to note that each of these sub-categories contains a huge number of activities. For example, there are many different sports, and many ways to participate in sports-related hobbies, and the same is true for all other categories listed. There are also other hobbies which do not obviously fall into any of these categories; examples include juggling or trick cycling. Given such a heterogeneous field of activities, it is difficult to identify common elements that distinguish the term 'hobby' from the term 'activity.' However, Stebbins claims that serious leisure is distinguished by the following six characteristics:

- The occasional need to persevere.
- Leisure careers shaped by individual contingencies, turning points, and stages of achievement.
- Significant personal effort using specially acquired knowledge, training, experience, or skill.
- Durable benefits in the form of personal and social awards.
    - Personal rewards: personal enrichment, self-actualization, self-expression, self-image, self-gratification, recreation after work, financial return.
    - Social rewards: social attraction, group accomplishment, contribution to group maintenance and development.
- A unique ethos (community spirit) emerging for each serious leisure activity.
- Participants in serious leisure tend to identify strongly with their chosen pursuits.

However, for the purposes of this discussion, I consolidate these characteristics down to two that are particularly relevant. First, hobbies are intrinsically enjoyable for hobbyists. If this were not true, they would quit. Second, hobbies require skill, and hobbyists build skill by participating in the hobbies. If interactive systems are designed so that these elements are present (i.e. the systems require skill, build skill, and are intrinsically enjoyable) then it stands to reason that these systems have the potential to be used as hobbies; that is, some people would use these systems frequently over longer periods of time. The resulting hobby-system would be an example of a type of interactive media I call 'practice-oriented' HCI.

Practice-oriented interfaces are designed to draw volitional usage in order to facilitate the practicing of an activity. The achievement of long-term goals, the surmounting of long-term challenges, and the building of expertise all require regular effort over the course of long periods of time. Practice-oriented interfaces have a unique potential to help users accomplish these things. Fields such as education, physical fitness, rehabilitation, and productivity could benefit greatly from practice-oriented HCI. For example, in some applications, complex and rapid-paced interactions are required from users. These situations are like music performance in that practice is required. Interfaces can be designed to facilitate that practice and make it enjoyable. Similarly, sometimes large-scale complex deliverables are needed from users. Obviously, practice-oriented interfaces might help users stay on task; but perhaps less obviously, practice-oriented interfaces might also help users attain skill at producing the deliverables more quickly and easily. An analogy illustrating this lies in the difference between composers and improvisers of music. Whereas a composer might painstakingly score musical events over the course of hours or days, a practiced improviser might create music of the same complexity with little effort, taking only the requisite time to produce the notes on the instrument. Practice-oriented HCI may prove to be an important concept as new tangible, gestural, and motion analysis-based interfaces emerge in the marketplace. Such interfaces often afford a greater degree of nuance and technical skill than traditional keyboard-and-mouse interfaces,

but nuance and technical skill may only be attained through practice. If people practiced these interfaces in the way amateur musicians practice instruments, they might eventually be capable of efficient, nuanced, and technically-skilled interface control. Another implication of practice-oriented HCI is that interfaces can be designed for the sole purpose of getting users to focus more awareness on their own bodily movements. Movement awareness is beneficial in itself—ergonomic movement techniques such as the Alexander Technique are built around it (Jones 1997).

One ultimate goal of my research is to provide a framework helping designers create effective practice-oriented HCI, so the interfaces they design will draw the largest voluntary user bases, and the most volitional usage per user, over the longest time frames possible. One of the main approaches I took in order to accomplish this task was to explore the field of hobby, by examining many hobbies in order to identify common elements that impact long-term volitional participation. In addition to the serious leisure literature, introspection on my own experiences as a hobbyist was useful in this effort. My own hobbies tend to be based in music. For example, I am a maker and player of folk instrumentation, as well as a composer and songwriter. Therefore, this dissertation puts a heavy emphasis on music. All three thesis projects contain an element of musical interaction, and much of my design framework was informed by my experience of engagement with musical activities.

It is worth noting, however, that musical hobbies were not chosen merely because of my personal bias as a musician—as discussed in Wallis et al. (2013), they are especially well-suited as templates for practice-oriented HCI. For one thing, music is one of the more prevalent hobbies in existence. According to data from the U.S. Census Bureau[1], 4.2% of U.S. citizens play a musical instrument one or more times per week. The only hobby-like activities with greater incidence of regular participation were food-related (e.g. baking, cooking), puzzles (e.g. sudoku, crossword), and playing cards. From a research perspective, however, focusing on

---

[1] see www.census.gov/prod/2011pubs/12statab/arts.pdf

music is more advantageous than focusing on these other hobbies, because there is a great deal of literature authored by music educators on strategies for engaging new musicians (e.g. Hallam, 2002). Also, instrument playing entails the use of musical devices. This makes it especially relevant to the field of HCI, which is similarly concerned with the usage of devices. Just as human-computer interfaces are designed to simplify and enable a variety of complex tasks, instruments exist to simplify and enable the act of music generation.

Hobbies based on musical instruments illuminate some of the potential benefits of practice-oriented HCI. Musical instrument hobbyists accrue certain benefits from their regular participation. If practice-oriented HCI could be developed which drew participation in the same way that instrumental hobbies do, these benefits could be leveraged to purposes that are not necessarily musical. For example:

- Instrument practice is a way for musicians to gradually attain their long-term goals of musical expertise. Interfaces could be similarly based on the attainment of long-term goals.

- Instrumentalists practice in order to attain skill so they can perform complex music more easily. This is a useful paradigm in HCI when difficulty cannot be avoided: practice-oriented HCI might facilitate the attainment of skill thereby allowing users to manage higher levels of difficulty.

- Instrument learning results in nuanced and masterful bodily movement in instrumentalists. Tangible, gestural, or motion-based interfaces also frequently require nuanced and masterful movement; so if they are designed to elicit frequent practice, they may be more successful (cf. Leman, 2007).

Anecdotal evidence suggests that instrument playing is a large and multifaceted hobby where each musician takes a unique perspective on the hobby. Some musicians especially enjoy

the social aspects of their hobby. Others enjoy instrument collecting, making an effort to find instruments with desired tonal or cosmetic qualities. Many musicians enjoy making their own instruments, and many are happy simply exploring the sound, kinesthetics, and mental states that arise from practicing in solitude.

2.1.2   Activity Design as a Form of Art

In this dissertation, I contend that activity design can and should be considered its own art form: an art form with an aesthetic based on participation and engagement. This may seem out of context as an art form because aesthetic philosophy in the past has centered on art forms with perceptually tangible results such as musical performances or paintings. Activity design produces results that are less tangible: the medium is the audience's participation in itself. For example, if I invent a card game which becomes popular, the activity I have designed has no physical form and is not wholly perceptible to the senses. It is a meme held in the minds of the participants, existing only on a conceptual level. I contend that there is still an aesthetic judgement to be made about this activity. However, in other forms of art, aesthetic judgements refer to beauty; but in activity design aesthetic judgements should refer to how enjoyable or engaging the activity is. Therefore, understanding and discussing the art form of activity design requires modification of certain aesthetic concepts such as medium, genre, function, and form. Accordingly, I propose modifications to the definitions of these aesthetic words in the following paragraphs:

Medium: In artistic contexts and in other contexts such as human-computer interface design, the term medium is used to denote means by which communication takes place. In traditional forms of art, however, this generally means the materials by which the artwork is made. Therefore, the aesthetic beauty of a painting might be communicated through the medium of oil on canvas, where the aesthetic beauty of a musical piece might be communicated via the

18

medium of piano and violin. This concept of artistic medium began to change, expanding to include less perceptually tangible sorts of artistic "materials," with the introduction of interaction and new media art. For example, in an important treatise by Myron Krueger on artistic responsive environments, he has this to say (Krueger, 1977, pg.430):

> The environments described suggest a new art medium based on a commitment to real-time interaction between men and machines. The medium is comprised of sensing, display and control systems. It accepts inputs from or about the participant and then outputs in a way he can recognize as corresponding to his behavior. The relationship between inputs and outputs is arbitrary and variable, allowing the artist to intervene between the participant's action and the results perceived. Thus, for example, the participant's physical movement can cause sounds or his voice can be used to navigate a computer defined visual space. It is the composition of these relationships between action and response that is important. The beauty of the visual and aural response is secondary. Response is the medium!

Based on this logic, if activity is the artwork, participation is the medium. However, this definition of medium, by itself, lacks specificity: essentially the same, or very similar, descriptions of medium would broadly apply to many otherwise incomparable activities. Therefore, I propose that medium in activity design should refer also to the materials or hardware that participants must use in order to participate. Therefore, the medium of my hypothetical card-game, referenced above, would include the deck of cards needed for participants to play.

Genre: Typically the genre of an artwork marks it as belonging to a category of the art-form that engenders a distinct form or style. In activity design, genres will be similar to already-existing activities, denoted as (for some examples): game-like activities, puzzle-like activities, musical activities, and so forth.

Function: Typically, the function of an artwork refers to any reasons the artist had for creating a particular piece. Examples include internally-based functions such as a desire for self-expression or a desire to be creative; they also include externally-based functions such as a desire to inspire cultural enlightenment or political change. Most if not all artistic functions

available to other forms of art will also be available in the art-form of activity design. However, there are some functions which may be uniquely applicable to the art-form of activity design. Two of these are:

- The inspiration of engagement and flow among the participants: Engagement is beneficial to people. Scientists such as Csikszentmihalyi (1991) believe flow and engagement to be one key to life-long happiness. Bringing about these feelings in activity participants is a worthy endeavor.

- The advancement of individual participant skills or, more broadly, the advancement of specific cultural qualities. Participating in an activity results in skill growth in that activity, so the design of new activities can help participants attain desired outcomes in terms of certain skills. Extended to the broader population of a large user base, this fact results in audience-wide advancement in skill at certain tasks. For example, one function of creating an engaging new sport may be increased physical fitness among the participants of that sport.

In this artistic context, function refers only to an artwork's *raison d'être* from the viewpoint of its creator. This differs in other contexts, such as user interface design, where 'function' is often synonymous with 'application.' To obviate confusion in this dissertation, when discussing the usage of interfaces or systems I will use the latter term. Thus, practice-oriented systems have their respective applications, but share the trait of affording activities, and those activities have functions that are meaningful to the designer.

Form: One school of aesthetic philosophy, aesthetic formalism, holds that an artwork's value can be judged based on qualities which can be observed by the senses. Given this, it must be true that an artwork's worth or lack thereof is somewhat separate from the viewer's subjective experience; beauty is not merely in the eye of the beholder. The fact that many people

agree on the beauty of certain artworks indicates that that there are specific commonalities of that artwork which many people find aesthetically charged. Those commonalities are formal properties. Examples of such formal properties are descriptors such as smoothness, tone, color, mood, and so forth. However, applicable formal properties vary greatly between genre, medium, and especially art-form. For example, most of the formal properties of music are not applicable in the field of sculpture. Even if a formally descriptive term such as smoothness applies to both art-forms, the meaning is dramatically changed.

A description of an artwork's formal properties is not sufficient to describe the artwork itself. If it were otherwise, we could measure the aesthetic value of a painting merely by measuring observable qualities such as hue, density, lines, brush-strokes, and so forth. However, paintings can not be judged, or understood, unless they are seen—because each painting is different than the sum of its formal properties. Any formal property's impact on the overall aesthetic value is unknowable, because how the properties combine to make a work of art is unknowable. The combinatorial possibilities between formal properties are effectively infinite. Each person interprets this combination differently, and there are differences in interpretation even within each formal property.

The formal properties of the art-form of activity design will not be listed here. Chapter 3 delivers these in the form of heuristics. As will be discussed in section 2.2, there are many conceptual similarities between heuristics and formal properties. The heuristics given in chapter 3 are desirable properties of afforded activities when designers are creating engaging, participation-drawing practice-oriented interfaces.

2.2 Design and HCI

At some level, this dissertation addresses a problem underlying many societal issues: lack of willpower. This is not a problem anyone can fully solve; it has many traits in common with the intractable class of problems called "wicked problems" which are defined in Rittel and Webber (1973). True solutions to these problems can never be found, only better or worse ones, because all solutions involve the balancing of tradeoffs. Wicked problems are often difficult because they involve factions of stakeholders with widely divergent viewpoints and incentives. The willpower problem differs: instead of opposing factions, willpower is made intractable by opposing impulses within each individual. Most agree that physical fitness provides tremendous health and social benefits—there is no real opposing faction. Yet health problems due to lack of exercise persist in society, because each person is fighting an internal battle, where one impulse desires greater physical fitness, and another impulse desires to conserve effort and time.

The willpower problem, like wicked problems, is resistant to formal modeling, and this makes engineering approaches unworkable. Zimmerman et al. (2007) advocates that wicked problems be addressed by designers instead, because designers bring three processes that are useful for these sorts of problems (pg. 4):

- Designers bring "a process for engaging massively under-constrained problems that are difficult for traditional engineering approaches to address."

- Designers bring "empathy for users as part of the process. In addition to considering their needs and desires from an external-observer's perspective, designers work to embody the people they make things for."

- Designers bring "a process for integrating ideas from art, design, science, and engineering, in an attempt to make aesthetically functional interfaces."

22

The willpower problem is massively under-constrained and resistant to engineering approaches, and influencing the internal battles of individuals demands an approach based on empathy, and my natural way of researching involves integrating ideas from various disciplines; therefore, I have adopted many of the approaches outlined by Zimmerman. As stated in Zimmerman et al. (2010), there are several points where design and research intersect. One of these points is research for design (pg. 313): "Outcomes of [research for design] include frameworks, philosophies, design recommendations, design methods, and design implications." Since this dissertation is primarily focused on creating a design framework, it clearly falls under the category of research for design. On the other hand, seeing that designing actual systems was the vehicle for conceptualizing this design framework, there is also a strong element of research *through* design, which is another point of intersection in Zimmerman et al. (2010). Research through design depends on designing artifacts (i.e. prototypes, systems, interfaces, or other concrete designs) that suggest, illustrate, or embody a potential future. The artifacts designed in the process of research through design are manifestations of initial design theories, and can serve as exemplars that help illuminate and spread the underlying theories. Research through design can be combined with research for design, as the act of designing can help ground conceptual frameworks and guiding philosophies in praxis, and this is the approach I take in this dissertation.

There are parallels between the act of design and certain qualitative research methods. For example, action research (AR) methodology, which has been used most frequently in the context of problem-solving through policy change in institutional contexts, is a paradigm where researchers participate in problem-solving through making institutional policy changes, then documenting the results of those changes. According to most reviews of research methods in HCI (e.g. Myers, 1997; Kjeldskov and Graham, 2003), action research methods have been largely overlooked in the field. This may be beginning to change, because, as Hayes (2011, pg. 7) puts it: "...the cyclic approaches common to HCI, including those from user-centered design,

are similar to AR in privileging iteration and building on past experiences." Both action research and iterative design depend on a cycle where the researcher or designer makes some change, notes any response to that change, then plans more changes to address the new situation. In essence, the act of iterative design, assuming that the responses to new iterations are well-documented, is a form of scientific research.

In this section there are discussions of dissertation foundations and relevant concepts taken from the field of human-computer interface (HCI) design. These discussions contextualize important aspects of this dissertation's goals. Section 2.2.1 clarifies the concept of frameworks, thereby defining a template for the ultimate objective of this dissertation, the design framework on volitional usage. Section 2.2.2 defines experiential media, thereby describing what the design framework is applicable to. Section 2.2.3 places this dissertation in the context of existing research by discussing the work of other HCI researchers pursuing the development of frameworks, applicable to interactive media, that have topics that are closely related to volitional usage.

2.2.1 Frameworks

Thus far in this dissertation, there has been frequent use of the term 'framework,' but no explanation of what frameworks are. This subsection exists to remedy this lack of explanation. Here I will cover what frameworks are, then provide examples of existing frameworks that are relevant to my work. In the context of this dissertation, frameworks are conceptual models that can be useful in organizing the thoughts of designers. Perhaps they are defined most succinctly in Edelson (2002, pg. 114): "A design framework is a collection of coherent guidelines for a particular class of design challenge." Frameworks are inherently advisory, not dictatorial. They are meant to be used in the process of designing systems, but this would not be possible if they could not also be used to analyze and think critically about existing systems—if a framework is not useful in identifying areas of potential improvement in existing designs, then it can

not be useful in creating superior designs. Edelson equates his conception of frameworks with the conception of design principles given by Van den Akker (1999), who provides an explicit definition. According to Van den Akker, design principles make statements of the following nature (pg. 9): "If you want to design intervention X [for the purpose/function Y in context Z], then you are best advised to give that intervention the characteristics A, B, and C [substantive emphasis], and to do that via procedures K, L, and M [procedural emphasis], because of arguments P, Q, and R." This definition of frameworks helps delineate another term which has been frequently used in this dissertation without definition: heuristics. Using Van den Akker's formulations, heuristics can be considered the advisable characteristics, letters A through C. Thus, heuristics are useful as aides in the description of systems, and by extension, they are also useful when designing systems.

Consider an abstract case where a designer is faced with a challenging design task, and the reason the task is challenging is because the design must address some issue that has complex and unpredictable variables. One way to approach such a task is to divide the problem into smaller pieces, by identifying some general design qualities that: a) are relevant to the problem at hand, b) suggest what the final design should look like, and c) are easy to characterize, conceptualize, and use. If this approach were taken, these general design qualities could be used as rules-of-thumb. These rules-of-thumb are heuristics, and the set of heuristics—along with justifications, implications, interrelationships, and other discussion—is the framework. The design framework in chapter 3 follows this paradigm.

The usefulness of heuristics is not limited to systems design; they also facilitate the assessment and critique of existing systems. In this way, heuristics are similar to aesthetic formal properties as discussed in section 2.1. Just as aesthetic formal properties are useful descriptors when discussing art, heuristics are useful when discussing designs. However, there are limits to the descriptive power of both formal properties and heuristics. Just as an artwork's formal properties offer only part of the answer to the question "What is this piece of art?", heuristics

offer only part of the answer to the questions at the heart of their frameworks. The experience of an interactive system depends partially upon the heuristics of its design, but also partially upon the perceptions of the beholder, so it is impossible to use heuristics to make exact claims regarding how individuals might experience a system. As with formal properties, it is unlikely that knowledge of a system's composition, in terms of its design heuristics, can allow one to make quantitative overarching statements about it. The sum of a system's heuristics do not constitute the whole, so even in cases where heuristics are measurable, the measurements will not provide an accurate picture of that system's overall operation with regard to the framework. In many cases, heuristics and the interrelationships between heuristics are complex and hard to measure, and impossible to compare in an apples-to-apples way across systems. This seems to be the case for the heuristics given in chapter 3, for example.

One framework I use as an exemplar is the set of usability heuristics outlined in Nielsen and Molich (1990). As will be discussed in section 3.1, the quality of usability is highly important for volitional usage, because usable interfaces result in fewer distractions and frustrations, thereby allowing flow and engagement in users. The heuristics are summarized briefly as:

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Help users recognize, diagnose, and recover from errors
- Help and documentation

These heuristics match all the constraints given in the abstract case above. They clearly address the problem of usability, they provide concrete suggestions for design, and they are easy to characterize, conceptualize, and use. According to Nielsen and Molich, if designers incorporate these qualities into interfaces, the results will probably be more usable. No claim is made that these heuristics can stand alone: the complexities of design mean that considerations which are outside the framework's scope must be taken into account in order for a design to be successful.

Tensions may emerge between heuristics during the design process. Using the usability framework as an example, it is easy to envision a scenario where user control and freedom conflicts with error prevention. Resolving such tensions is possible because tradeoffs are allowable between heuristics. Heuristics are not to be treated as strict rules, rather, they should be treated as loose guidelines, mental shortcuts, or catalysts for thought. There are two reasons for this. First, heuristics may not be based on data-driven evidence alone; they are often based upon personal experience and analytical reasoning as well. Therefore, they should not be treated as though they are law. Second, in order for heuristics to be relevant in a wide range of design situations, they are usually defined somewhat vaguely. Heuristics outline generally desirable characteristics, but the designer must determine the details of how those characteristics will be incorporated because each design situation will demand a different manifestation of the heuristic.

2.2.2 *Experiential Media*

Frameworks are often useful in that they serve to illuminate and define the topics they are applied to. Often the framework addressing a design challenge provides the best available definition of that design challenge. For example, the term 'experiential media,' which is central to my dissertation, is hard to define but has an informative framework. Experiential media is a relatively new and unique development in the field of interactive media, discussed heavily in

Rikakis et al. (2006), Sundaram and Rikakis (2006), and Chen et al. (2009). The most succinct definition of experiential media might be "interactive media designed to cause users to have experiences." Unfortunately, this definition lacks specificity because it describes all interactive technology. To describe experiential media in a more useful way, it is necessary to delineate the concept of 'experiences.' The seminal aesthetics text "Art as Experience" (Dewey, 1934) states that experiences have the following characteristics:

- Experiences have beginnings and endings. In this way, individual experiences differ from experience as a whole, which is continuous.

- Experiences have unity. Any episodes between the beginning and end of an experience fuse into a single whole.

- Each experience is unique, with its own plot, emotion, rhythms, and other internal factors resulting in some unique pervading quality. If one event is indistinguishable from another event, neither can be considered an experience.

- Experiences are emotional. Emotions are inseparable from experience; they are entangled with events and objects, and arise automatically from any movement or change. No experience can occur without some emotional quality.

- Experiences have a structure, and that structure can be expressed as follows: a person undergoes something then, consequently, is affected by some outcome of undergoing that thing. This, in turn, causes the person to respond by doing something else, and the process continues in a cycle until some sort of equilibrium or harmony is achieved. At this point of equilibrium the experience can be said to have come to a close.

If these five characteristics were used as heuristics for designing new interactive systems—i.e. if interaction was designed to be unique, have unity, elicit emotion, and so forth—the resulting systems should provide an aesthetic experience per Dewey's definition. Such a system

can be called experiential media. The last characteristic listed is especially noteworthy: a design implementing the cyclical process of action, feedback, and reaction should contain the other four characteristics as well. According to Dewey's concept of experience, they will naturally emerge. Therefore, experiential media offers a cyclical type of interaction where the system responds to the user and the user in turn responds to the system, ad infinitum. This feedback loop between user and system means that each session, users can take a unique path to the point of equilibrium. It also dovetails nicely with the feedback loop between mind and action implied by activity theory (as discussed in section 2.1). For this reason, one might postulate that experiential media can be apt for facilitating growth within users.

The recent emergence of experiential media is driven by advances in technology. As expressed in Rikakis et al. (2006, pg. 13): "We are ... witnessing a rapid decline in the cost of sensing, storage, computing, and display. Thus sensors (audio, video, pressure, tangible), computing, ambient visual and sound displays and other feedback devices (vibration, light, heat) can now be co-located in the same physical environment creating a real-time feedback loop. This allows for development of a rich contextual understanding of human activity, at different scales of time and space, and offers the possibility to affect human activity in a radically new way." In older forms of media, production was separated from consumption; e.g. films were made, then distributed to audiences. In experiential media systems, production and consumption happen at the same time and place, and are often performed by the same people.

According to Sundaram and Rikakis (2006), designers of experiential media systems require knowledge in five domains. The first three domains have to do with technical components. They are:

- **Sensing:** Knowledge of technologies and techniques allowing an experiential media system to detect the physical world and human activity.

- **Feedback:** Knowledge of technologies and techniques that allow the experiential media system to communicate to humans using the various modalities.

- **Perception and Modeling:** Understanding of human perception, cognition, and behavior; plus knowledge of technologies and techniques for modeling these things, so that the experiential media system can be designed to be appropriately coupled to the users.
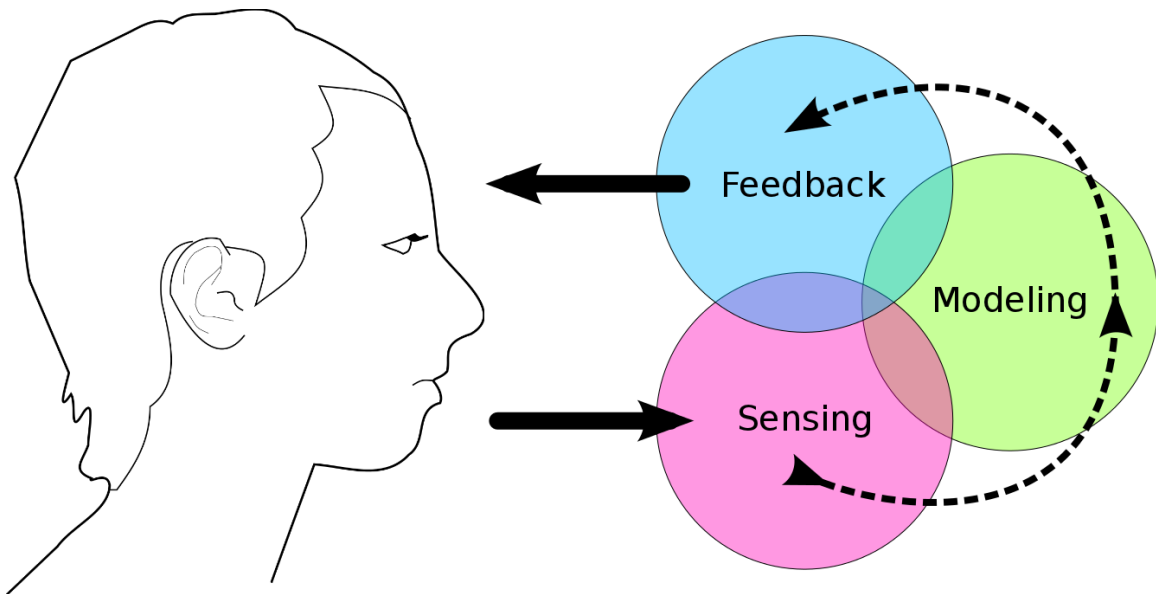
Figure 2.2: Flow of information in an experiential media system.

Figure 2.2 relates these three knowledge domains to the feedback loop in an experiential system. The sensing, modeling, and feedback circles overlap to show that, despite the conceptual separability of these three components, they are not always easily separable in actual implemented designs. For example, in many systems the modeling is inextricably built into the sensing component, the feedback component, or both. In addition to these three domains of knowledge, there also exist two other domains that Sundaram and Rikakis deem important for experiential media design. They are:

- **Experiential Construction:** Knowledge of how to combine sensing, modeling, and feedback to create a fully manifested experiential media system; plus knowledge of how a system design stems from its goals and intended function.

- **Learning and Knowledge:** Knowledge of how to evaluate the resulting media system, identify weaknesses and make iterative improvements to the system; plus knowledge of documentation and discussion regarding system construction and system goals.

These five knowledge domains can be treated as heuristics; they can help organize thought while designing experiential system, and they can act as guidelines on what to discuss when discussing experiential systems. I will frequently refer to them in chapters 4 and 5 when describing experiential system designs. The research in this dissertation will inform knowledge in all five domains, but is especially relevant to the intersection between experiential construction and perception and modeling.

2.2.3 Existing and Related Research

Thus far in this section, I have examined existing work that informs, but does not significantly overlap with, the contributions of this dissertation. Overlapping research does exist, however, in domains such as gamification, musical instrument design, and user experience (UX). To contextualize my dissertation, it is necessary to discuss similarities and differences of approach between my work and this existing work. For example, there are clear areas of overlap between the topic of volitional usage and the topic of user engagement: if volitional usage is to continue for any length of time, users must be engaged. For this reason, at one point in the history of this dissertation, I considered user engagement to be its main topic. However, engagement was always treated as a means to the end of drawing volitional usage, so eventually I modified the topic accordingly.

In the field of user engagement, work by O'Brien and Toms (2008, 2009) and Attfield et al. (2011) has the most similarity with mine because their work also contributes frameworks and heuristics. O'Brien and Toms (2008) looks at theories of play, theories of flow, and treatises on aesthetics to identify ten potential attributes of engagement: aesthetic appeal, attention, challenge, endurability, feedback, interactivity, user control, pleasure, sensory appeal, and novelty. These attributes were treated as interview guides during numerous informal interviews on the users of an online shopping interface, a search engine, a webcasting application, and a video game. The data from these interviews was coded according to the four 'threads of experience' discussed in McCarthy and Wright (2004): the thread of emotion (i.e. emotional connotations of the experience), the spatio-temporal thread (i.e. time, place, setting of the experience), the sensual thread (i.e. observable qualities of the experience), and the thread of composition (i.e. the structure of the experience; how the experience goes together; the beginning, middle, and end). This data analysis was then used to conceptualize a model of user engagement. In this model, engagement is an experience consisting of a beginning called the 'point of engagement,' a middle called the 'period of engagement,' and an end called 'disengagement.' If a user disengages but becomes engaged again within the same usage session, this is called 're-engagement.' As users go through the usage session and the process of engagement, they experience the different attributes of engagement at varying points of the process and at varying levels of intensity.

O'Brien and Toms (2009) builds upon the theory from O'Brien and Toms (2008) to create the Engagement Scale, a way of rating user engagement using a self-reported survey questionnaire. This survey questionnaire was tested on users of an online shopping interface. Based on the results of that test, the number of attributes of engagement was reduced to six: perceived usability, aesthetics, novelty, felt involvement, focused attention, and endurability. Also, a predictive model was developed that shows the relative importance of each of these attributes with regard to the overall engagement level reported by the users. The six attributes

32

all have a measure of independence, but noteworthy relationships exist between certain pairs of attributes. For example, highly-rated aesthetics are predictive of perceived usability in this model.

In a position paper, Attfield et al. (2011) suggests that the attributes of engagement should be focused attention, positive affect, aesthetics, endurability, novelty, richness and control, user context; and lastly, reputation, trust, and expectation. Attfield also surveys existing means of evaluating user engagement. The Engagement Scale is held forth as an exemplar of subjective evaluative methods, although Attfield notes that O'Brien and Toms' goal of creating a scale of engagement generalizable to all HCI is complicated by the fact that "User engagement almost certainly has different characteristics in different application domains and for different demographic groups" (pg. 4). As evaluations based on self-reporting have weaknesses due to their subjective nature, Attfield also discusses evaluation techniques that are considered more objective. Two of these objective assessment strategies were applicable only to web evaluations, but the others–subjective reporting of time, follow-on task performance, and physiological sensing–bear discussion here:

- Subjective duration assessment: Studies show people tend to overestimate time spent on challenging activities and underestimate time spent on challenging activities (Baldauf et al., 2009; Czerwinski et al., 2001). It is possible that this fact could be leveraged in evaluations of user engagement. More research is needed before this evaluation strategy is verified, however, because the link between challenge and engagement is not direct. High challenge could be indicative of usability problems (as Czerwinski postulates) or it could elicit a feeling of flow (as discussed in section 2.3).

- Follow-on task performance: A study discussed in Jennett et al. (2008) suggests that if people are engaged in an activity, but then told to interrupt that activity to perform

another task, they will tend to perform that second task more slowly. This fact could be leveraged in evaluations of user engagement.

- Physiological measurement: This strategy encompasses the use of a wide range of sensors, such as eye trackers, cameras, and biosensors to evaluate affective state, focus of attention, heart rate, and many other bodily factors which pertain to engagement.

Although the frameworks of O'Brien and Toms (2008, 2009) and Attfield et al. (2011) are useful for the purpose of understanding engagement, I believe that my framework is better suited for designers of new practice-oriented interfaces, because my framework differs in the following three ways: First, my framework is based on the process of design, rather than the process of engagement; therefore it encapsulates only those aspects of engagement that the designer has some sort of control over. Second, my framework views engagement as a driver of usage, rather being an end in itself; this is more relevant to systems that require long-term usage such as practice-oriented systems. Lastly, because my framework is based on the idea that interaction is a vehicle for activity, it can and does integrate ideas from other engaging activities, such as hobbies, while these other frameworks are based only on specific interactive systems. In the following paragraphs I will expound on these areas of difference.

Unlike my framework, which is meant to be a tool for designers making *new* practice-oriented systems, the work in O'Brien and Toms (2008, 2009) and Attfield et al. (2011) is focused on understanding user engagement in existing systems that were designed by others. For this reason, these frameworks include characteristics that are internal to the user, where my heuristics focus exclusively on system qualities that are under the control of the system designer. For example, Attfield's attributes of engagement describe states held by engaged users, such as focused attention and endurability; my heuristics are concerned with how to elicit these states. Attfield et al. (2011) says "Our goal is to define a framework in which user engagement can be studied, measured, and explained, and, as an ultimate aim, lead to recommendations

34

and guidelines for user interface and interaction design for front-end web technology" (pg. 6). O'Brien and Toms take a similar approach. I reverse this approach by conceptualizing the recommendations and guidelines, in the form of heuristics, first. Later I evaluate them by using them as analytical lenses in the course of iterative design and evaluation.

O'Brien and Toms (2008, 2009) and Attfield et al. (2011) provide strategies for evaluating engagement. Although these evaluation methods seem useful from a research standpoint, they are problematic from a design standpoint because in many cases they would be impractical for designers to use on their own systems. These evaluation methods require carefully controlling the conditions and environment of the evaluation. This means that the experience of system evaluators will differ from the experience of typical system users. Testers will need payment or reward to participate in the tests; it is unlikely they will feel intrinsically motivated to wear sensors or fill out surveys. According to self-determination theory (as discussed in section 2.3), such incentives will negatively impact motivation. All of these methods of evaluation are intrusive to some degree and will change the user engagement levels which they purport to measure.

In my work, the role of engagement is to drive usage; this differs from the engagement-centric perspective of O'Brien and Toms (2008, 2009) and Attfield et al. (2011). For example, both frameworks incorporate an attribute of endurability, which can be defined as "positive impressions of a system after usage, accompanied by a desire to re-engage in usage." In O'Brien and Toms (2008, 2009) and Attfield et al. (2011), endurability is one indicator of engagement. In my work, it is the reason engagement is desirable. The model of engagement discussed in O'Brien and Toms (2008) has structural similarities with a model of volitional usage I diagram in chapter 3 (fig. 3.1). However, the O'Brien and Toms model is concerned with the interior peaks and valleys of engagement within each usage session, while mine is concerned only with the usage sessions themselves. This has impacts upon terminology. In their model, disengagement and re-engagement happen many times during the course of every usage session. Although

fluctuations of engagement undoubtedly exist within every usage session, in this dissertation the terms 'disengagement' and 're-engagement' refer only to points when users actually quit or restart system usage.

Another way my dissertation differs from the work in O'Brien and Toms (2008, 2009) and Attfield et al. (2011) stems from the fact that I consider interaction to be a vehicle for activity (as discussed in section 2.1). Treating interaction this way allowed me to take the approach of looking at non-mediated activities (for example, hobbies such as instrumental music-making) as sources of inspiration for experiential systems design. This, in turn, let me synthesize and integrate theories of motivation, such as self-determination theory (discussed in section 2.3), when constructing my design framework. Exploring these fields has offered a wealth of insight into engagement and led to framework heuristics (e.g. ownership, cooperation, demonstrability) that are not addressed as attributes of engagement in the frameworks under discussion.

Like the field of user engagement, the fields of serious gaming and gamification have aspects that overlap with the work in this dissertation. Work such as Malone (1982) is particularly similar because it offers heuristics, derived from the world of video gaming, outlining how to design enjoyable user interfaces. Malone's main heuristics are challenge, fantasy, and curiosity. There are areas of overlap between these concepts and some of the heuristics in chapter 3, but there are also areas of difference, stemming from the fact that this dissertation was written in the hope of moving toward practice-oriented interfaces not based entirely on gaming. Although gamification is a valid approach to creating practice-oriented interfaces, and many existing practice-oriented interfaces are game-like, this approach is unlikely to be successful for some users who do not find video games compelling or whose practice-oriented goals cannot be addressed via video games. Therefore, the heuristics in chapter 3 are based on an expanded view on the topic of volitional usage, one that encompasses gaming but also includes other activities that draw volitional usage, such as hobbies. Malone's heuristic of fantasy—dealing with the ways in which games offer compelling narratives and metaphors—is a particularly

36

clear area of difference. It implies that interfaces designed for enjoyment should contain elements of story; where my activity-based framework acknowledges that both story-based and non-story-based activities can elicit enjoyment and draw volitional usage.

There is a category of interface design that is focused on the design of new musical interfaces. The principal venue for scholarship in this realm is the conference on New Interfaces for Musical Expression (NIME). However, after exhaustively searching the proceedings of that conference from its inception in 2001, I have determined that my goals and the goals of NIME are largely orthogonal. Although some NIME publications are concerned with borrowing ideas from HCI in order to improve the process of creating new musical interfaces (cf. Wanderley and Orio, 2002), none were found that were concerned, as my work is, with applying lessons from music-making (and other hobbies) more broadly to the realm of general interaction design. The most similar work found was Johnston (2011), which advocates applying grounded theory methodology to the evaluation of new musical interfaces. This is ultimately the route I chose to take when evaluating the system described in chapter 5.

The field of user experience (UX) also has overlapping facets with the field of volitional usage, as much UX literature is focused on the creation of fun and enjoyable interfaces (cf. Blythe et al., 2004; Hassenzahl and Tractinsky, 2006). Hassenzahl's focus on hedonic and ludic interfaces makes his work especially relevant. For example, Hassenzahl (2008) is, thus far, the only paper I have come across that explores HCI design through the lens of self-determination theory, which is an approach I take (as discussed in section 2.3 and chapter 3). The work in this dissertation could be viewed as moving toward the goal of improved user experience, except in my work the real end goal is volitional usage, not just positive user experience. Positive user experience is an important factor in volitional usage, but it is not the only factor. This has led me to include heuristics into my framework that have little impact on what people might experience when using a system, but much impact on whether they choose to try using the system in the first place.

In this subsection, I have outlined the differences between my work and the existing work of various other researchers in different scholarly domains. It is important to note that none of the existing work under discussion is being criticized here. On the contrary, the research discussed here is highly useful; it is just not quite as focused as this dissertation is on the narrow topic of how to design practice-oriented interfaces, or how to design experiential media for volitional usage. Still, in chapter 3, it will be apparent that some of the conclusions I reached are similar to conclusions reached by these existing researchers.

2.3 Emotion, Engagement, and Motivation of Users

This dissertation is based on the idea that some activities, by dint of their intrinsic qualities, persuade people to participate, and that these intrinsic qualities can be integrated into experiential systems to encourage potential users. This premise is fundamentally dependent on the psychology of potential users; therefore, in this section I discuss concepts from psychological theories of emotion, engagement, and motivation. These concepts have been incorporated into my design framework in various ways, and they have informed the design process in a number of the interactive systems I have had a significant role in developing; for example, I make frequent reference to many of these concepts when discussing the design of systems in chapters 4 and 5.

2.3.1 Emotion

There is something paradoxical about the stated goal of the design framework in chapter 3: it is based on the idea of encouraging voluntary behavior. If an action is voluntary, it is done out of intrinsic desire; the word 'voluntary' implies that no encouragement should be needed. Encouraging volitional usage, then, is a process that operates on the 'desire' part of the problem; that is, the framework is designed to get users to *want* to volunteer. This requires transforming users on an emotional level. Looking at hobbyists, one can see that their emotions drive them to participate—as discussed in Stebbins (1982), hobbyists are often highly passionate about their pursuits. For example, Stebbins encountered people who claimed, in all seriousness, to love their hobby more than their significant others. If Dewey (1934) is correct and every experience is distinguished by its own emotional character, then it stands to reason that we volunteer to participate in activities because of the feelings they elicit.

Therefore, in this subsection, I will discuss emotion from the standpoint of media design. Before I can do that, it is necessary to explain how I use the term 'emotion.' Some

psychologists delimit the word to short-term affective states such as jealousy or joy (e.g. Scherer, 2000). In this view, other mental states such as enjoyment or desire would be be considered attitudes instead of emotions. I take a broader view, using a definition for emotion found in the *Oxford English Dictionary*: "Feelings derived from one's circumstances, mood, or relationships to others" (Soanes and Stevenson, 2006). This allows me to treat certain mental states that are important to HCI, such as engagement and motivation, as under the purview of affective design.

Psychological models of emotion are useful for designers because they help organize, categorize, and explain emotions. One simple yet scientifically validated emotional model is the Circumplex Model (Russell, 1980), a two-dimensional model where the y-axis corresponds to emotional arousal or intensity and the x-axis corresponds to emotional pleasure or valence. As seen in fig. 2.3, common emotion words, when rated along these axes, fall into a ring around the origin with words like anger in the upper left, joy in the upper right, serenity in the lower right, and sadness in the lower left. The wide agreement with regard to the valence and arousal of each emotion supports the idea that mood is universally understandable.

The circumplex model is intuitive and easy to use in interfaces (for an example, see fig. 4.14), but it is not entirely emotionally accurate. For instance, anger and fear are close to one another on the circumplex, but most people perceive these as highly different emotions; more different than other closely-spaced emotions such as sadness and depression. The only way to overcome this drawback is to add more dimensions to the model, such as a dominance axis which separates anger and fear (e.g. Russell and Mehrabian, 1977). A plethora of multidimensional models of emotion have been developed using statistics and factorial analysis. Some of these models have as many as twelve dimensions.

There are two types of emotional design challenges that designers face. The first is designing to present emotions that can be perceived by the users or consumers. The second is designing media that induces felt emotions in the users or consumers. In many ways, the

A CIRCUMPLEX MODEL OF AFFECT
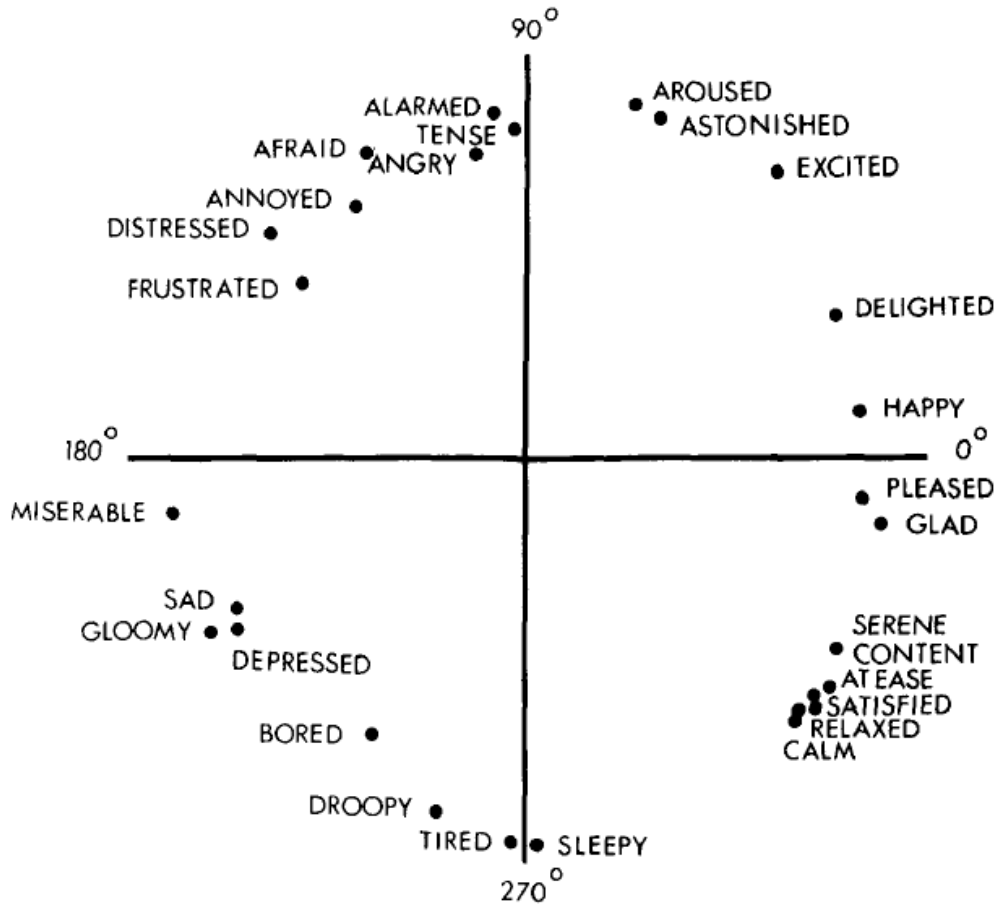
Figure 2.3: Circular scaling of mean ($n = 36$) rated arousal-valence coordinates of 28 affect words. From "A Circumplex Model of Affect" by James A. Russell, *J. Personality and Social Psychology*, 1980, 39, 1161-1178. Copyright 1980 by the American Psychological Association. Reprinted by permission.

first design challenge is easier, so I will start my discussion with that, then later I will discuss approaches for emotional inductions with regard to a few relevant emotional states. For clarity, from this point on, when discussing emotion that is perceived externally (as opposed to felt internally) I will use the term 'mood.'

Designing media that has a perceptible mood is made possible by the fact that emotion is akin to a broadly-accepted form of communication that can be understood via the mediums of: sight, as when reading facial expressions; sound, as when reading tone-of-voice; or touch, as when feeling the vibrations of a purring cat. Moods are recognized and agreed upon by most who perceive them, although counterexamples exist—some moods are ambiguous and can be interpreted in slightly different ways, such as the mood of the facial expression of the Mona Lisa. There seem to be some differences between cultures with regard to the perception of facial expressions and tone-of-voice, but these differences are specific to certain emotions, so there are also many intercultural similarities (cf. Ekman and Friesen, 1971; Elfenbein and Ambady, 2002). Despite these counterexamples, according to Elfenbein and Ambady (2003, pg. 160): "The communication of emotion has a strong universal component. For example, people of different cultures can watch foreign films and understand much of their original feeling. Likewise, people can develop strong bonds with pets while communicating largely through nonverbal displays of emotion."

Mood is unique for its contagious property; it often seems to communicate directly to our subconscious. For example, simply being in the presence of cheerful people can often make us feel happy ourselves (cf. Barsade, 2002). This contagious aspect of mood is one reason why mood is such an effective medium for communication. Mood seems to leverage the part of our mind that handles empathy, treating it as a channel of information. When we perceive emotion in an agent, whether that agent be a person, a pet, a movie, an object, or a piece of music, what we are really doing is performing a structural analysis on that agent. Some component of the structural analysis answers the question "What emotions would drive me, personally, to exhibit

these structural features?" If we hear someone singing sweetly we assign a pleasant emotion to that music, because it would take a pleasant emotion to drive us to sing that way ourselves. When we hear fast-paced music we assign it the same high-energy emotions we might feel if we were to move our bodies at a fast pace, or watch a movie with a fast-paced editing style, or even drive a car at high speed. Speed is an example of something I call a 'universal emotional feature'—a structural feature observable in many situations and mediums that carries a similar emotional connotation whenever it is observed. Other examples of these include harshness and amplitude. Harshness is observable in language, in tone of voice, in color schemes, and many other mediums; and in all its forms it carries a similar relationship with emotional valence. Amplitude is observable in bodily movement, in brush strokes, in some status symbols, and many other things; and in all its forms it carries similar emotional connotations with regard to confidence or dominance.

There is a theory of emotional perception and emotional contagion which is proprioceptive in nature, positing that certain structural features observed in outside agents seem emotional because they correspond with movement features which are emotional for evolutionary reasons. For example, at some point in human history, odds of survival increased if high-arousal emotions such as anger or terror correlated with rapid rates of movement. Therefore, if we observe agents moving rapidly, even if they only do so in the conceptual sense (e.g. high-tempo music), it can seem as though they have high-arousal emotions such as anger or terror (Chen et al., 2008; Ekman et al., 1983). Similarly, if we observe these same agents moving slowly, it can seem as though they have low-arousal emotions such as serenity or sadness. Embodied cognition theory (e.g. Niedenthal, 2007) and mirror neuron theory (e.g. Freedberg and Gallese, 2007) support this proprioceptive view. This view proffers an interesting approach to designers of practice-oriented HCI applied to movement training and rehabilitation: through affecting the emotions of users, an interface can also affect the movements of users (and vice versa).

In section 2.1 I discussed music as a form of activity. Here in this section I will discuss music as a form of media, because music is rich in examples for designers who want to portray or impact emotion. It is one of the most frequently-used ways of setting mood, and a great deal of literature has been written that links music structural features to the moods they engender (for a review see Gabrielsson and Lindström, 2001). Interestingly, music is not good for portraying all emotions; complex emotions such as jealousy or nostalgia require objects of focus (i.e. something to be jealous of, or something to be nostalgic about). Music by itself cannot provide these objects of focus, but if the objects are simultaneously provided through other means such as lyrics or imagery, these moods can be portrayed (Collier, 2002).

In addition to being an effective medium for mood, music also provides many everyday examples of emotional induction. One of the ways emotion can be musically induced is through emotional contagion: sometimes listeners absorb the musical mood and begin to feel it internally (cf. Juslin and Vastfjall, 2008). This seems to take place more frequently when music accompanies other emotional information. For example, background music is a key factor in how audiences identify, on an emotional level, with characters in films (Marshall and Cohen, 1988). From this, it can be observed that the emotional quality of music can color an experience even when listeners are not consciously aware of the music (i.e. when they have mentally tuned the music out, as many film watchers do). Perhaps this ability for musical mood to penetrate distracted minds is related to the proprioceptive theory of emotional contagion described above. Numerous models of musical induced emotion have been developed. Perhaps the most rigorous of these is the Geneva Emotional Music Scale (GEMS) developed by Zentner et al. (2008), which has nine dimensions—wonder, transcendence, tenderness, nostalgia, peacefulness, power, joyful activation, tension, and sadness—which can be factored down to three dimensions: sublimity, vitality, and unease. Many people have physiological emotional responses such as chills, goosebumps, or racing hearts in response to specific musical moments. However, not everyone does, and although some studies have identified features of the

44

musical moments which are more likely to cause such strong feelings (Sloboda, 1991), the moments exhibiting these features do so unreliably. The common thread running through these musically-induced emotions that are marked by physiological responses is that they seem to occur on moments of musical surprise. Music is well-suited to induce the emotion of surprise because it operates from within a well-established set of expectations, and when expectations are violated surprise is the result (cf. Meyer, 1956).

Music is also known to commonly induce a feeling of resolution. Resolution is based around the idea that humans prefer relaxed states to tense states, so moving from tension to relaxation provides a moment of relief. For a non-musical example, imagine being present while a cup of water, carelessly placed, topples on the edge of a table. Nobody in the room could relax until the glass either falls and spills or (preferably) is moved to a safer location. The relief felt when either of those things happens is an example of resolution. Musical harmonies provide the same feeling of relief in a less dramatic way, by moving from chords of high tension (i.e. discordant chords) to chords of low tension (i.e. stable, harmonious chords). Musical tension and relaxation is discussed at length in Huron (2006, ch.15).

As stated previously, designing to induce emotion is a more challenging task than designing to present moods. There are several reasons for this. Emotions are induced through highly complex processes that are not fully understood at this point. Different people react differently, on an emotional level, to the same stimuli. And lastly, each specific emotion a designer might want to induce requires an entirely different approach. For example, inducing surprise requires setting up expectations, then violating them, and inducing resolution or relief requires moving from an unstable state to a stable state. Whereas, a fundamentally different method is required for inducing calmness, which might best be approached by creating a calm atmosphere in the hopes of emotional contagion. As the designers of Microsoft's Clippy avatar can attest, it is easier to induce emotions on accident (such as emotions of annoyance toward the designer) than it is to induce emotions intentionally (Swartz, 2003).

For this dissertation, the most useful emotional states to induce are related to engagement and motivation. These emotional states have well-established psychological theories. In this subsection I will discuss engagement, along with two related concepts, immersion and flow. Understanding of engagement is best approached from a standpoint of information transfer, and this requires an explanation of the concept of 'modalities.' In human-computer interfaces (HCI), modalities are mediums of communication between interface and human. Obrenovic and Starcevic (2004) provides a simple definition for modalities; in their models, "an HCI modality engages human capabilities to produce an effect on users" (pg. 66). Modalities are often thought of as directional channels that allow information to be transmitted from a source to a receiver. Therefore, interfaces have sensing modalities that allow humans to interact with them (e.g. keyboards for text, cameras for optical tracking), while humans have modalities allowing interfaces to communicate back. Here, I am concerned mainly with the human modalities. With humans, it is important to distinguish between sensory modalities and representational modalities.

The most frequently-employed sensory modalities are sound, vision, touch, or some combination of these; although rare, olfaction and taste interfaces also exist. The modality of touch can be thought of as an umbrella of several other senses having to do with bodily motion, such as proprioception (the sense of how one moves in relation to one's own body) and kinetics (the sense of how one's body relates to the surrounding world). Generally speaking, when the term 'multimodal interfaces' is used, this means that multiple sensory modalities are used in combination. Of course, this needs to be narrowed because almost all interfaces could be considered to be multimodal in some way. Even a device such as a CD player could be considered a multimodal interface, because it uses buttons that require haptics and kinetics. Such a broad definition is not useful for this dissertation, so here I delimit multimodal interfaces

to include only interfaces using multiple sensory modalities in a continuous and coordinated way. This excludes the CD player, but includes musical instruments.

Considering that all interaction performed by humans requires either motion or the senses, one could make the case that sensory modalities are the only true modalities. However, the taxonomy of modalities given in Bernsen (1994) also includes representational modalities: modalities transmitting representational forms of information such as images or language. Representational forms of information are important to this dissertation topic, so I use a broad definition for modalities. In my view, if information can be transmitted using a single self-contained method, then received and understood as intended, then its means of transmission can be considered a modality. Based on my earlier formulation of mood as a near-universal way to communicate, mood is one example of a modality. This shows that some representational modalities can utilize multiple sensory modalities. That is, mood can be transmitted through hearing (e.g. musical mood, tone of voice) and vision (e.g. emotional imagery, facial expressions); other representational modalities can use multiple senses as well. Music is a representational modality which cannot—it uses only sound. Interactive music built into systems for the purpose of transmitting information to users is called 'musical sonification' (discussed heavily in section 4.1). This shows that representational modalities can carry other representational modalities. That is, music can convey emotional information (i.e. on the modality of mood) while conveying symbolic information (via sonification); this is possible with other representational modalities as well. Many other abstract concepts, such as metaphor and gesture, also can be categorized as representational modalities. Clearly, these representational modalities can be interrelated and congruent; musical metaphors and emotional gestures are evidence of this.

Designers who wish to use a modality should understand it thoroughly first in order to ensure the information being transmitted is appropriate; it might not make sense to try to deliver spatial location information using music if one has the option to use some visually-based

modality, for example. Bernsen (1994) may be a useful resource in identifying the best interactive mappings for a modality. Assuming designers do use the modalities to transmit appropriate information, however, there still remain other factors that will determine the legibility of the communication. In an effort to describe these factors, I propose a perspective of modalities called the "channel model." In the sense that modalities are methods of communication, they can be likened to speech. Likening modalities to speech affords an explanation of modalities borrowed from information theory, as information theory has frequently been applied to language. In this explanation, modalities, like speech, are modeled as transmission channels of limited bandwidth. They are essentially separate but parallel channels; this explains why speech and music can simultaneously convey different information even though they both depend on the sense of hearing. It is important to note that this channel model is proposed only as a useful abstraction, as it is not based on neurological research.

The channel model implies certain rules or heuristics; therefore it can be thought of as a framework for designing multimodal communications. The proposed rules are:

- **Moderation, i.e. "talking too fast":** Too much information conveyed too quickly overwhelms. Imagine listening to a recording of instructions sped up to twice or three times the normal rate. Most likely, you could not understand, remember, or follow such instructions. This overloading of the speech modality is possible, and should be avoided, in all modalities. If we say, arbitrarily, that a certain communication channel can support 100 pieces of information per second, then it stands to reason that the modality will be overloaded if you give it 150 pieces of information per second. At least 50 pieces of information will be lost, but probably more because some of the lost pieces would be important for context, linking concepts together in the overall sequence. To illustrate this using the above example of the sped-up recording, even if the lost parts of speech were all taken from the end of a sentence, the first part of any truncated words would be hard to decipher, and the meaning of the entire sentence could be lost. This guideline

also has an interpretation—based on the concept of density—for static modalities. For example, in a visual layout or an image, if important information is packed too densely in space, the result will not be legible.

- **Coherence, i.e. "multiple unrelated conversations":** Confusion results if communication on a modality is not cohesive. Imagine listening to a broken radio that automatically switches between different talk channels every word or two. No meaning could be extracted from the spoken words. Now imagine listening to two radios simultaneously that are set to different talk channels. It is unlikely you could understand or retain much of what either host says. From an information theory perspective, these two experiences are effectively the same. In a transmission channel, all incoming information goes through the same bottleneck regardless of source or context. Therefore, in the example of the two simultaneous radios, the incoming data would get intermingled and the resulting stream of words would be just as meaningless as the words played by the broken channel-switching radio. It might be possible for a person to understand the words coming from one of the radios, but only by completely ignoring everything from the other radio. Another way of looking at this guideline is as if the bandwidth of the communication channel is always set too low to be able to handle more than one stream of information at a time; presumably, two streams of information would require double the bandwidth of one. However, if the different streams of information are heavily related to one another, such that they can be fused into a single stream of information (e.g. as in a conversation with multiple people where everyone is on the same topic and most wait their turn to speak), then the channel might be able to handle the bandwidth. As in the previous rule, an interpretation exists for this rule allowing it to be applied to static media. For example, a visual layout with multiple elements, where the elements

are either: a) completely unrelated to one another, or b) placed in such a way that they obscure one another, will not be legible.

- **Concision, i.e. "inane or boring conversation":** This can be alternatively thought of as a rule against redundance. Boredom or exasperation results if communication is too repetitive or uninformative. Imagine listening to a broken record endlessly repeating the same phrase or two. Alternatively, imagine listening to someone speak endlessly about inane or trivial things. Often we would rather sit in silence than do one of these things. In this channel model, using a modality takes some effort. If only a small amount of information were being transmitted on that bandwidth, leaving most of the bandwidth unused—or, if the information on the bandwidth were obviously compressible to a small fraction of the bandwidth, as in the case of the broken record—the benefits of using that modality would not be worth the effort. This would explain why we might "tune out" the modality entirely, or, if this is not possible, become increasingly annoyed with that aspect of the interaction. This is analogous to other modalities: for example, highly repetitive music can become annoying to video game players even when the game is otherwise engaging.

- **Synchronization, i.e. "speech and body language together":** Information is reinforced or emphasized when given in several ways at once. For example, a good speaker will impart information not only using words, but with body language, facial expressions, tone of voice, and props or visual aids. If each modality is separate, then it is possible to present unrelated or contrasting information in different modalities. However, when two or more modalities present the same or similar information, the result is an emphasis of that information. From the standpoint of my channel model, where using modalities requires some effort, it is logical to deduce that the synchronized information is important, because it would be more efficient to transmit that information using only one modality,

but effort is being expended to transmit it in duplicate. From an information theoretical standpoint, this also indicates that the meaning of the information will be more clear, as noise may corrupt one copy of the information while the other copy remains intact. Multimodal interfaces can emphasize information by communicating on multiple modes simultaneously. Similarly, as long as the information does not conflict too much, more streams of information can be disseminated if they are transmitted using separate modes.

These rules can result in complex multimodal designs because tensions and interrelationships arise between them. For example, the rule of coherence suggests that separate streams of information should be transmitted in separate modalities, but if one of the streams contains little information, this conflicts with the rule of concision. For another example, violating the rule of concision may be allowable if synchronization is also used, because it is possible to envision a scenario where the information complexity arises between modalities instead of within them. A repetitive musical gesture might be allowable, for example, if users control when the gesture will be triggered. Or, a repetitive gesture might be allowable if triggered whenever an animated character performs a certain action, even if that character is not controlled by the user.

As it happens, this conceptualization of bandwidth-limited channels of information is applicable elsewhere in the field of psychology, most notably in the realms of cognitive load and engagement. In this field the bandwidth can be mathematically estimated, as shown in Csikszentmihalyi (1991), where the estimation is used to provide a neurological explanation for the theory of flow. His explanation can be expressed as follows: According to Miller (1956), we can hold about seven, plus or minus two, pieces of information at a single time in our working memory. The compression strategy of chunking—memorizing several pieces of information as a single unit—allows our working memory to contain longer sequences of information. For example, we can hold phone numbers in the working memory because the area codes are treated as chunks of information rather than separate digits. According to Orme (1969) we

become aware of new information at a rate of about eighteen times per second. Multiplying the capacity of the working memory by eighteen gives us a bandwidth of somewhere between 90 and 162 chunks of information per second. More information than this cannot be cognitively processed.

LeCompte (1999) argues that because Miller's experimentation was based on the limits of working memory, designers should base designs on a working memory with a smaller capacity (three pieces of information) so that information will be presented in a way that can be easily processed. Given this working memory of three, the bandwidth would be 54 chunks per second. Even basic everyday tasks such as listening to speech demand a great deal of cognitive attention. For example, Csikszentmihalyi, by treating phonemes as chunks of information, estimates that listening to speech takes around forty chunks per second. Research by Campana et al. (2011) suggests that the context of speech frequently allows people to predict entire words based on initial phonemes; this affords chunks consisting of multiple phonemes, which would mean that Csikszentmihalyi's estimates could be revised downwards in some cases. Still, these bandwidths indicate that it is unlikely that most people can easily attend to multiple unrelated streams of information, as discussed in the rule of incoherence above. This is an argument for minimizing distraction in interface designs.

However, in the context of a single complex experience, the using up of the entire attentional bandwidth is frequently beneficial, inciting feelings of immersion or flow. Immersion and flow might be considered two sides of the same coin. In immersion, the entire cognitive bandwidth is used by absorbing information through the passive modalities (modalities not related to movement). One place where immersion is frequently experienced is in movie theaters. In flow, the entire cognitive bandwidth is used by performing some challenging task. Some of the cognitive bandwidth is still used to process information through the senses, but only the information which is useful for task accomplishment is attended to. Some activities which frequently incite flow are sports, musical performance, or video gaming. Immersion and flow

take place within the context of single sessions, and are considered highly enjoyable by those who are experiencing them. They are experiences to be pursued. People experiencing immersion or flow have reported forgetting where they are, losing track of time, forgetting about physical concerns such as hunger, or even losing their sense of self for a time. Remembering these things requires cognitive bandwidth, and all the bandwidth is used in the immersive or flow-inducing experience. In the flow state, people often also report being able to achieve previously unachieved levels of performance.

2.3.3 Motivation

The rest of this section is devoted to theories of motivation. These are important to this dissertation for two reasons. First, they explain how people select between different activities; i.e. why someone might choose to take up *this* hobby when they could just as easily take up some other one. Second, they explain why people choose to have hobbies at all, and why they persist in them for such long periods of time. The motivational theory most central to this dissertation, self-determination theory (SDT), is good at explaining hobby persistence. SDT applies equally well to all forms of serious leisure, however, so it does not go far enough in explaining hobby selection. Therefore, I will discuss other theories first, as they provide more insight on that question. Although the theories under discussion have separate provenances, there is a great deal of agreement between them.

Maslow (1943) provides a tiered theory of motivation, where physical needs such as food and shelter are lower-level and thus take precedence over other types of motivation, such as love or self-actualization. Maslow's hierarchy of needs is frequently diagrammed in the form of a pyramid (see fig. 2.4) to represent the fact that people will focus on unmet needs lower on the pyramid to the exclusion of unmet needs higher on the pyramid: e.g. when people feel hungry or unsafe, they do not worry about issues regarding friendships, confidence levels, or

creativity. People will only shift focus to higher levels of the pyramid once needs are met at the lower levels.
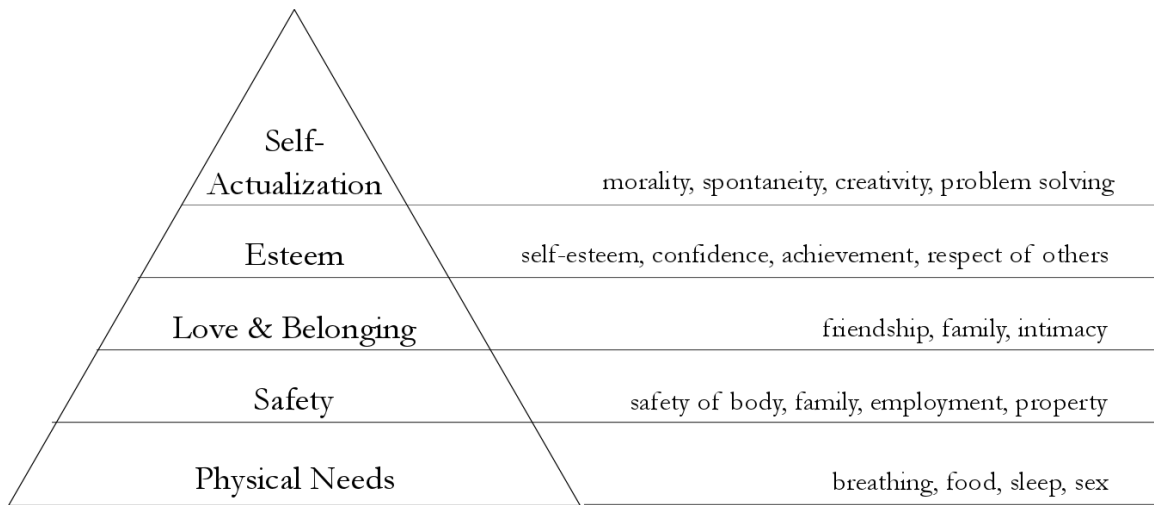
Figure 2.4: Maslow's hierarchy of needs.

Erikson and Erikson (1998) provides a theory of identity that is age-specific; it details how identity changes with age. In this theory, each life-stage is associated with a driving impulse. The life-stages and the driving impulses are given in the list below. The driving impulses may influence the sort of activities people choose to take up. For example, the elderly may prefer activities with nostalgic components. Younger adults may prefer activities containing social components. High levels of exploration or competition might be incorporated into activities designed for kids. The life-stages and driving impulses that are relevant in this dissertation are as follows:

- Initiative (3–6 years): Driven to do things for oneself.
- Competence (6–11 years): Driven to compete, acquire abilities and skill.
- Identity (12–20 years): Driven to explore and express oneself.
- Intimacy (20–34 years): Driven to date, marry, make friends.
- Generativity (35–64 years): Driven to positively impact family, society, the world.
- Ego Integrity (65 and on): Driven to look back with a sense of accomplishment.

Reiss (2004) posits that sixteen desires motivate all behaviors in humans and animals. Differences in behavior are caused by different levels of importance placed on each basic desire. Reiss's sixteen desires can be used as a framework for identifying appealing elements of activities, explaining in part why some activities are chosen over others. If someone places a great deal of importance on the desire to eat, he or she might take up baking. If, instead, there is more importance placed on the desire to save, then he or she might enjoy collecting stamps, coins, or comic books. Hobbies based on playing team sports probably satisfy multiple desires including physical fitness, vengeance, order, and status. The sixteen desires are:

- Power: desire to influence
- Curiosity: desire for knowledge
- Independence: desire to be autonomous
- Status: desire for social standing
- Social contact: desire for peer companionship
- Vengeance: desire to get even, compete, win
- Honor: desire to obey a traditional moral code
- Idealism: desire to improve society
- Physical exercise: desire to exercise muscles
- Romance: desire for sex, including courting
- Family: desire to raise own children
- Order: desire to organize
- Eating: desire to eat
- Acceptance: desire for approval
- Tranquility: desire to avoid anxiety, fear
- Saving: desire to collect or be frugal

The theories discussed so far have points of agreement with self-determination theory (SDT). As discussed in Ryan and Deci (2000), SDT holds that motives for behavior are

55

either extrinsic (based on outside punishments or rewards) or intrinsic (based on personal enjoyment), and that there are three main intrinsic motives. These three motives will be discussed in greater detail in upcoming paragraphs, but to briefly summarize them: a) *mastery* is related to learning and skill attainment, b) *autonomy* is related to perceived freedom, and c) *purpose* is related to human connection. Therefore, SDT is compatible with Maslow's hierarchy of needs: the extrinsic motives are related to the two lower levels of the pyramid, and the intrinsic motives can be interpreted as an expansion on the three higher levels. Similarly, many of Erikson's driving impulses seem related to intrinsic motives. All of the adult life-stages seem focused on purpose, and all of the childhood life-stages seem related to autonomy or mastery. Reiss claims that his sixteen listed desires are incompatible with SDT because he considers them all to be intrinsic motives (i.e. he denies a split between extrinsic and intrinsic motives). However, in looking at the list of desires it is not difficult to identify relationships between them and different intrinsic and extrinsic motives. Independence is equivalent to autonomy; curiosity and physical exercise seem related to mastery; and power, status, social contact, vengeance, honor, idealism, romance, family, order, and acceptance all seem related to purpose. Eating, saving, and tranquility might best be thought of as driven by extrinsic motives.

Both Reiss and Maslow hold all motives to be similar in kind, meaning that an individual's motivation to do something is determined by adding up the impacts of the various needs or desires. SDT differs, because it considers motives to be either extrinsic or intrinsic. In empirical validations of SDT it was found that intrinsically motivated participants tend to perform better and persist longer in a given activity than extrinsically motivated participants. It was also found that incentivizing an activity extrinsically (using payment, for example) serves to decrease the level of intrinsic motivation of activity participants. Therefore, overall motivation to do an activity can be less than the sum of intrinsic and extrinsic motivation for that activity. Hobbyists do not receive or require extrinsic motivation, so the intrinsic-extrinsic division makes SDT a useful theory for the purposes of this dissertation—it allows me to consider the

intrinsic motives that drive voluntary behavior (as in the case of hobbyists), without consider-

ing the less relevant extrinsic motives that are based on payment or reprisal. I use SDT as a

psychological basis for the longer-term heuristics of the design framework in chapter 3. I treat

the intrinsic motives as qualities of activities, interfaces, or interface designs. All of these can

be characterized according to the degree they facilitate the intrinsic motives, and this allowed

me to examine how each intrinsic motive might be made to exist in interfaces.

People tend to enjoy activities containing the intrinsic motives of mastery, autonomy,

and purpose. For example, all hobbies seem to be driven by one or more of these motives;

instrument-playing has all three. The following paragraphs discuss these intrinsic motives in

greater detail, and also discusses some implications with regard to HCI design.

Mastery   As stated previously, the flow state is entered when an activity takes up so much

of our cognitive bandwidth that there is no bandwidth left over to give us our sense of self.

Therefore, performing an activity can only lead to flow if the inherent challenge of the activity

is matched to our skills in such a way that we can successfully perform the activity, but only by

devoting our utmost attention. If the activity presents too little challenge, we might experience

boredom or relaxation instead of flow, and if it presents too much, we might experience arousal

or anxiety.

Since practicing an activity leads to increased skill in that activity, in order to maintain a

sense of flow the challenge must be continually increased—but never by too much. It follows

from this that learning any complex activity such as instrumental performance will be a gradual

and repetitive process, in which one challenge is mastered, then an incrementally more difficult

challenge is pursued. Rank beginners need an activity with a low entry bar, where the entry bar

represents the amount of difficulty inherent in beginning the activity, but the length of time

the person is likely to remain interested depends upon that activity's ceiling, where the ceiling

represents an achievement pinnacle past which challenge can no longer be increased. Between
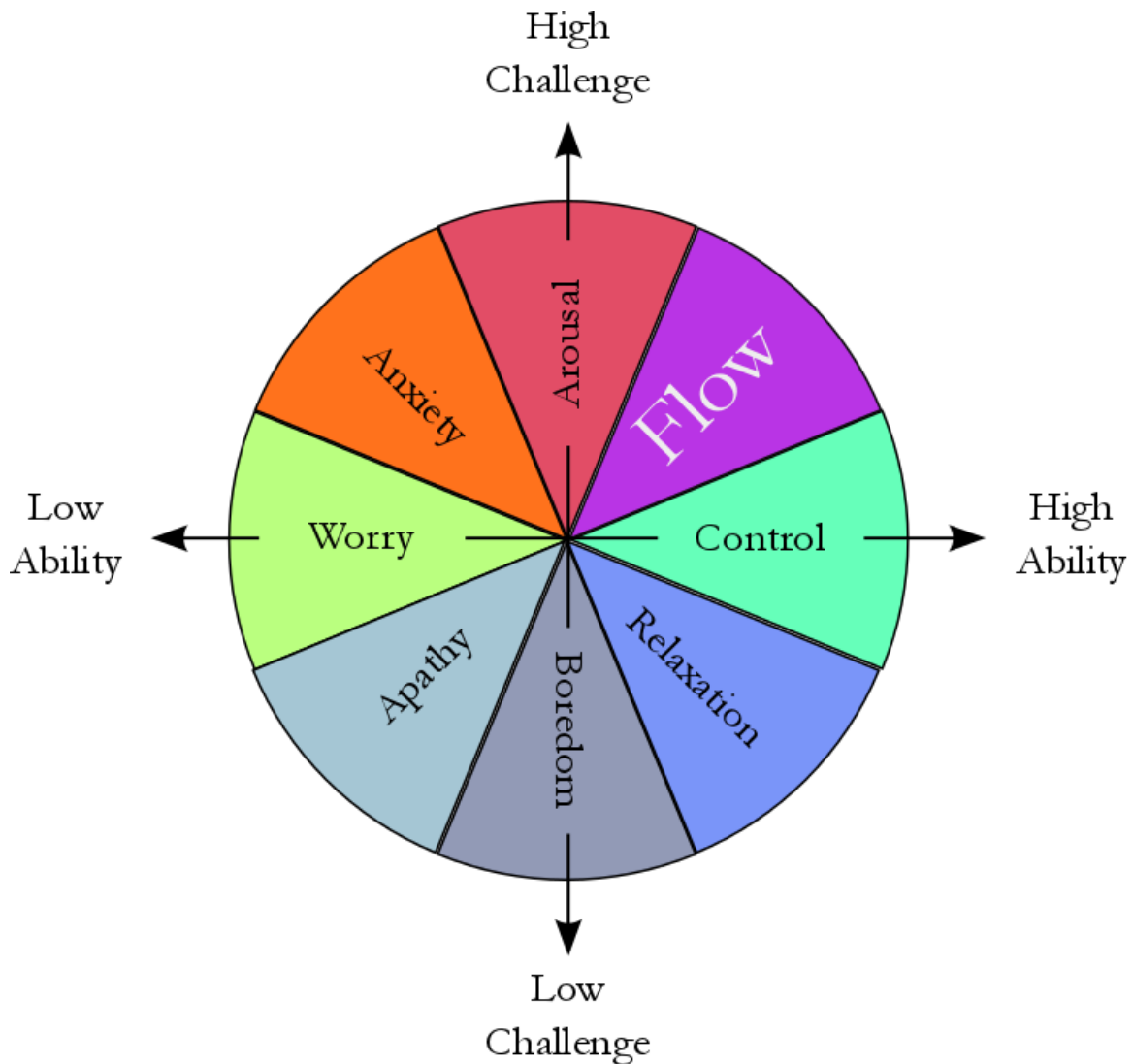
Figure 2.5: Emotional response to challenge level versus ability level.

the entry bar and the ceiling the difficulty must be increased smoothly and gradually. If it is not, plateaus in learning will result. Most athletes and musicians are familiar with the concept of plateaus. Plateaus cause people to get stuck at points in their progression toward mastery, sometimes for a long time—they can even result in a loss of interest in the activity. However, sometimes if a plateau is surmounted the learning will become gradual again.

Those who pursue an activity for the sheer joy of increasing their skills in that activity are said to be intrinsically motivated by mastery. In the world of interactive media, video games

might represent the best examples of well-managed mastery. They use scoring to help people evaluate their own skills, and use leveling to incrementally increase the challenge from easy tutorial levels to levels of extreme difficulty. The only area in which video games do not manage mastery well is in their ceilings, which are often too low. This makes sense from a business perspective: games will not be sold if everyone stays addicted to the same game for years on end. Looking to subscription-based video games such as World of Warcraft, however, we begin to see video games with ceilings approaching the height of hobbies.

Mastery is an important factor in the choice of activity. An activity will be more appealing to participants if they feel they can master it despite its difficulty. This means: a) people are more likely to choose hobbies if they have access to equipment, resources, or expertise that make mastery more likely; and b) people are more likely to choose hobbies if they believe they have aptitudes in the required skills. This can have ramifications for those who design activities for people with health problems or disabilities. For example, a person with a movement disorder may feel particularly discouraged about movement-based activities. In systems design, it is always important to show the target users that the system was designed with their specific issues in mind. Another way mastery factors into the choice of an activity is through inspiration. Watching experts at work, impressively demonstrating mastery in their skill, can inspire people to try and learn that skill and emulate that mastery.

Autonomy   Activities which we can engage in mastering are intrinsically enjoyable, but not so much when the impetus to do so comes from outside ourselves. Just because job-related activities can be mastered does not mean we enjoy these activities when our bosses tell us to do them. Activities are more intrinsically motivating if they contain a sense of mastery coupled with a sense of autonomy, or self-volition. Autonomous activities are self-selected or self-directed activities which do not seem mandatory. Participants choose to do these activities out

of the sheer desire to do them, and they also choose their own path to mastery within these activities.

Autonomy can be a challenging design goal for interactive systems. It is impossible to design a system that places no constraints on users whatsoever. Even if the physics involved with computerized sensing allowed this (and the physics do not), there are other practicalities involved. Each interactive system is designed for some application; that application can be considered a restriction on autonomy, and it is likely that some users would prefer some different one. Unless the system designer can create a wide range of varying systems for users to choose from, where each system has a different application, user autonomy is going to be restricted. However, limited options may not be a severe problem as long as there are at least some choices to be made by the user. Studies show that people often feel anxiety over a choice if they have too many options to pick from, or if they are given too much leeway in taking back previous choices (Gilbert, 2006). This fact seems to indicate that, although having great power of self-selection increases the sense of autonomy, there is such a thing as too much power of self-selection. This phenomenon is sometimes referred to as 'tyranny-of-choice' (Schwartz, 2004).

Returning to the interactive media example of video games, we see that, although video games have a set of rules which users are incapable of breaking, there is autonomy within those rules. Within the context of an activity, the term 'strategy' can be described as the intersection of autonomy and mastery, and a large part of mastering any video game is figuring out the best strategies. If multiple strategies exist which are equally advantageous to the player, the user will experience a greater sense of autonomy, which is why subscription-based video games such as World of Warcraft[2] are designed for a highly customized and configurable path through the levels.

---

[2]see http://us.battle.net/wow/en/

Purpose Mastery and autonomy are powerful intrinsic motives, and for some people they may be enough to motivate an activity over the course of a lifetime. However, at some level, most people learn a skill for its ability to help them feel connected with others. This sense of interrelatedness an activity imparts is its intrinsically motivating property of purpose. At first glance, it may seem that the motives of autonomy and purpose conflict with one another, because the quality of autonomy seems to appeal to our selfish or individualistic side. This is not the case, as studies show that those who report greater feelings of autonomy in their lives tend to actually be better at working in groups than others, and they also report higher feelings of interconnectedness with their peers, families, and society (Kim et al., 1998).

According to Stebbins (2007), the rewards of hobby participation include social rewards such as social attraction, group accomplishment, and contribution to group maintenance and development. I have found that discussion of purpose within hobby is facilitated by a terminology of sharing. Hobbies are shared in many ways; some obvious ways I have identified using anecdotal evidence are listed here:

- Demonstrating Skills: Even hobbies which seem to be highly solitary by nature are often partially motivated by the eventual sharing of the skill through demonstration. This might be the case with unicyclists or poets, who learn and practice their craft in solitude, but intend to show off their mastery.

- Competition: Some activities are competitive in nature, and the sharing component of these activities is the getting together for the purposes of competition. Examples of this abound, but some obvious ones include team sports and chess. The competitive aspect provides a way to evaluate and validate mastery in an activity.

- Communal Activity: Often hobbyists will get together simply to practice their hobby as a group. This is especially important to musicians who perform in ensembles, but groups exist for many other hobbyists, such as writers or painters, to workshop as well.

The community of hobbyists teach each other and motivate each other, improving the mastery within the community.

- Improving Society: This idealistic purpose is interesting because it adds to a participant's feeling of connectedness, not with individuals, but with people in the abstract. The hobbyist need not actually know any of the people he or she is connecting with. Artists and musicians may feel this type of purpose in their work, if they feel they are adding beauty to the lives of others and furthering the progression of art. Volunteers and activists of all kinds are also likely motivated by the purpose of improving society.

- Quality Time: In this type of sharing, the activity helps the participant connect or spend quality time with family and friends. An example of this might be a parent and adolescent who, as a hobby, work together to refurbish and customize an old automobile.

- Self-expression: Some artistic activities, such as music or painting, give participants a new way to express their thoughts, emotions, or creativity. This purpose includes many activities that result in improving a person's ability to communicate with others of to have access to others. For example, stroke patients might participate in rehabilitation activities in order to improve their ability to speak and get around, so that more opportunities to form connections will exist.

Purpose is also a highly important factor in the choosing of an activity. Some people are motivated more by the prospect of quality time with the family than by the prospect of improving society. Others prefer competition. People tend to choose activities that contain an element of purpose that matches their inclinations. If systems are designed so that purpose is added in multiple ways simultaneously (for example, a competitive system affording self-expression and the demonstration of skills), this may increase the chances that potential users will be interested.

Chapter 3

THE FRAMEWORK

In this chapter, I propose design heuristics that are meant to facilitate the creation of experiential media that draws volitional usage. These heuristics are divided into two sets: a set of heuristics for shorter-term volitional usage, and a set of heuristics for longer-term volitional usage. Because the ultimate goal of my research is to help designers create systems that draw volitional usage for long periods of time, the longer-term set of heuristics, given in section 3.2, was formulated first (Wallis et al., 2011b, 2013). As will be discussed in chapter 5, however, the process of research made it apparent that the longer-term heuristics are, by themselves, deficient. The reason should have been obvious: in order for longer-term volitional usage to take place, users must be attracted and engaged in the first place.

Obviously, participation in any hobby-like activity is not continuous over the course of weeks or months; rather, there are numerous repeated sessions of participation, where each session is freely sought by the participant. One way to think of these numerous sessions is as if they were each an example of short-term volitional usage. Therefore, systems designed for longer-term usage must necessarily also be designed for shorter-term usage. This realization led to an expansion of the goals in this dissertation: rather than focusing only on longer-term usage, I have now constructed a theory regarding the entire usage career, covering the time from when the user is first introduced to an experiential system to the time when the user sets that system aside never to take it up again. In this model, shorter-term engagement is distinct from longer-term engagement, and the early stages of a usage career can be viewed as a transition from the former to the latter. Sometimes, usage careers may end during this transition, especially if the system is not well-designed for longer-term volitional usage.

The following subsections discuss the shorter-term and longer-term heuristics separately. The framework has partial support, either from existing research or from data discussed in chapter 5, and all of the discussed heuristics contain an element of common sense, but it is important to note that at this point the framework can not be considered complete. Substantial future study and fine-tuning—more than can be undertaken within a single Ph.D. dissertation—will be required before the framework can be considered formal grounded theory. Still, at this time, these heuristics represent the best collection of analytical lenses for use when looking at or designing practice-oriented interfaces, as no other framework exists for designing experiential media for longer-term volitional usage.

Heuristic	Description	Impacts
Promotion	Users should be made aware of system's existence	Attraction
Convenience	System should be easily accessible to users	Attraction & endurability
Targeting	Match between mediated activity and intended user's predilections	Attraction & initial engagement
Novelty	System should offer a new experience to the user	Attraction & initial engagement
Cognitive Richness	Mediated activity should match user cognitive load	Engagement & endurability
Usability	System interface should minimize distractions from mediated activity.	Disengagement & endurability

Table 3.1: Heuristics for short-term engagement.

3.1 Shorter-term Usage

With any experiential system, if users continue to participate in more and more usage sessions, their reasons for doing so will naturally change over time. Many factors which might have attracted them initially are no longer the motivating factors causing them to continue re-engaging in session after session. In this section I address the factors which draw them initially. Given that short-term usage is necessary before longer-term usage can exist, how can we design systems that draw short-term usage? The cycle by which a person enters longer-term volitional usage with an ideal interactive system, if broken down into steps, might look something like the following:

1. Attraction: User desires to participate in session because of externalities and design features.

2. Initiation: User chooses to begin session.

3. Engagement: User is engaged during session, possibly experiencing flow.

4. Disengagement: User's session comes to a close, ideally accompanied by a feeling of satisfaction (not frustration or boredom).

5. Endurability: User is now attracted to system because of design features, but also because the system was previously enjoyable.

Step 5 is analogous to step 1, so at this point the cycle repeats. This cycle of engagement can be expressed as a state diagram as in fig. 3.1. Going by this model, the task of designing a system to draw first-time usage can be broken down into two parts. First, the system must be made attractive to potential users, and second, the system must be made engaging upon first-time usage so that people will be more apt to try the system another time. These two tasks are interrelated. For example, a system that seems like it will be engaging is more likely to attract first-time usage. Still, it is instructive to separately consider the factors going into each task.
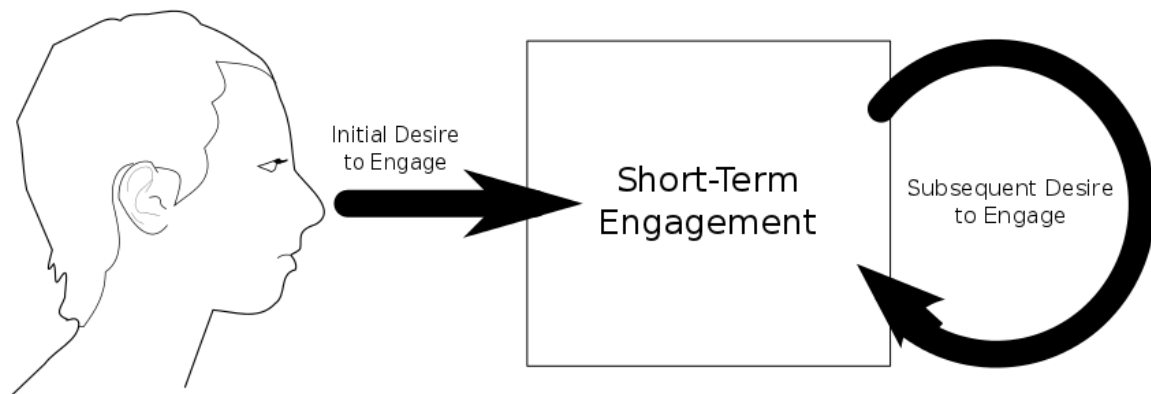
Figure 3.1: Repetitive cycle of volitional usage.

Attracting new users is a complex challenge, in part because it is vulnerable to factors outside the scope of the system design itself. For example, environmental externalities, settings, and demographics all play a large role. Consider a case where a designer hopes that some target user base will volunteer to try his or her system. Figure 3.2 gives a model depicting usage careers. In looking at fig. 3.2, we see that a large number of potential users is trimmed down, over time, to a small number of committed users. In the beginning of this process, many potential users never try a system because they are unaware of its existence. Of those who are aware, some will be incapable of using the system due to factors in their own lives; they have no time, or are concerned with more important things. Of those who remain, some will be simply uninterested in trying the system. Anyone who volunteers to try a system for the first time will

be aware, capable, and interested in trying the system. Therefore, attracting first-time usage can be expressed as a problem of minimizing the number of potential users who are unaware, incapable, or uninterested in using the system.
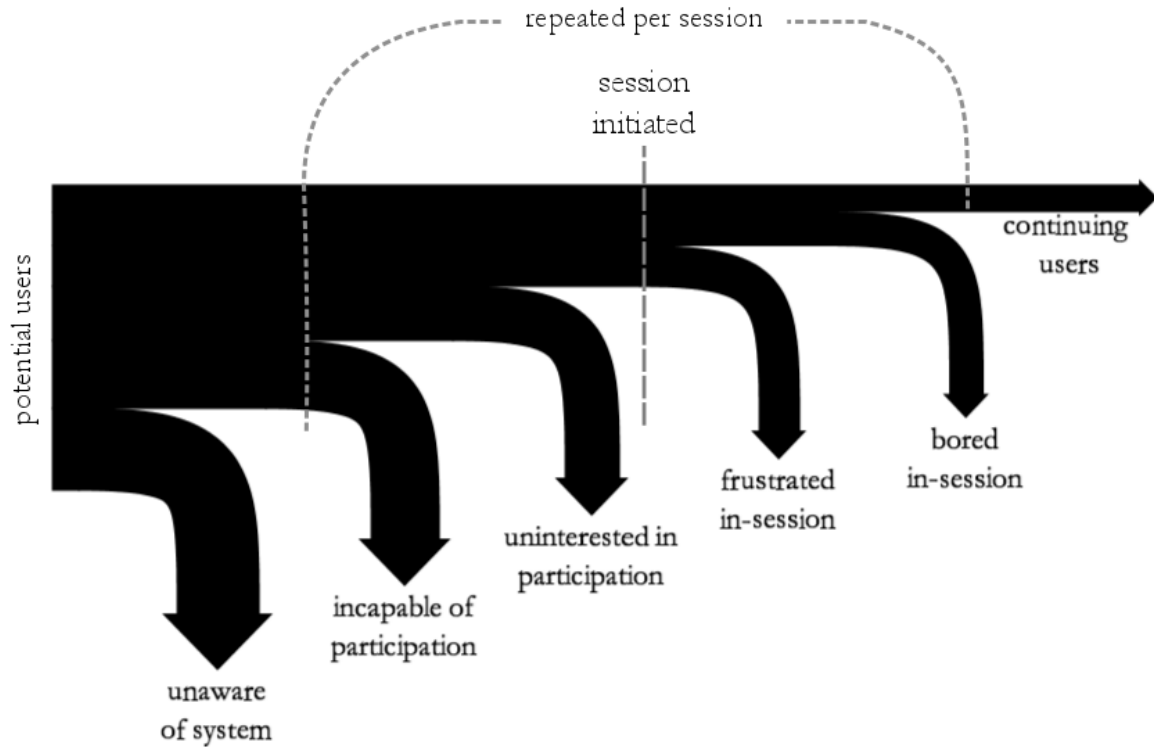
Figure 3.2: Model of user forfeit over time. Losses upon subsequent usage sessions will continue to decrease because remaining users will be more persistent.

Unaware users can be made aware, but this is a problem of marketing, not design. Marketing techniques are beyond the scope of this dissertation, but suffice it to say that *promotion* is an important variable in determining how many people initially try a system. Similarly, the number of incapable users can be reduced somewhat, because users with limited time availability may still try systems if those systems are made highly *convenient*. Convenience, like promotion, is not a factor which is purely within the realm of systems design. Often, convenience is a matter of managing the setting the system is situated in, the method of distribution, or other factors external to system design. Systems should be made easily accessible to potential users. This includes financial accessibility, so systems should be made cheaply if possible. In the case of

physical systems, systems should be located near potential users. Studies discussed in chapter 5 suggest that interactive systems will be tried more frequently when demonstrated at events or showcases where a target user base is congregated. This is due to convenience: the users are there, and they have spare time between talks during which they have nothing better to do. Similarly, if software-based systems can be easily and freely downloaded so users can try them in their spare time without leaving the house, more people who are aware of them might give them a try.

Only the third type of non-users, the uninterested, can be easily reduced in number solely through careful system design. One presumes that all uninterested users do have interests of some sort. Therefore, one way to minimize the number of uninterested is to identify any trends with regard to the interests of the target user base before commencing the design process, so the system application, goals, or afforded activity can be well-matched to the existing predilections of the desired users (i.e. *targeted* to those users).

This is the only aspect of the framework that implies an ordered design process that designers are advised to follow. A system design can stem from an idea-first process or a problem-first process. In idea-first, the designer has an idea for a system, so he or she searches for an application area for that system. In problem-first, the designer knows the application area first, because he or she is trying to solve some human-centered problem, and every aspect of the system (including the form and the afforded activity) is then designed to address that problem. Targeting system designs will be easier in a problem-first paradigm. This allows the designer to learn about the target user base before-hand, so that the afforded activity and form of the system can result from the process of reconciling the system's problem-based *raison d'être* with the likely interests of the user base. Systems drawing volitional usage are not strictly needed in the absence of human-centered problems. Many hobbies exist that already provide engaging experiences for people, so designing all-new hobbies is not called for unless

the new hobbies address a need. Also, systems will draw more volitional usage from busy users if benefits beyond engagement exist.

Targeting a design for a user group of any size is a tyranny-of-choice challenge. That is, there is an effectively infinite number of ways to address any human-centered problem, so picking a single design path can be difficult. One way to approach this task is to think of it as a process of ruling out options until a more manageable set is left. Asking the following questions may be helpful:

- **Looking at the user base, what existing activities are popular?** If a system is made for a certain demographic, are there any common elements among activities that are popular within that demographic? If there are any activity types that are popular amongst the group, can these be mediated to address the human-centered problem?

- **Looking at the psychology of the user base, what activities seem apt?** Various psychological theories may be helpful when targeting system designs. For instance, Reiss's sixteen intrinsic desires (discussed in section 2.3) may help identify activity types that the user base tends to find interesting. Case in point, some targeted groups may have more of an affinity for systems that are family-focused, while others may have more of an affinity for systems based on physical exercise. This question should be contrasted with the previous question, as any such affinities should be evident in the set of existing popular activities for the targeted user base.

- **What systems or activities am I, as designer, most interested and/or expert in?** A designer's own predilections can and should be taken into account. These will impact the quality of the system designs, and the quality of the designs will, in turn, impact user volitional usage and engagement. This question might be helpful in picking a final design path once the other questions have been used to narrow down the set of options.

Another factor of relevance to initial attraction, as well as to engagement during initial usage sessions, is the inherent *novelty* of the system. In some ways, novelty stands in opposition with targeting. That is, targeted systems dovetail with previous user interests, but novel systems are unfamiliar to users and seem unprecedented to them. These elements must be balanced, because a system that emulates an existing activity too closely risks seeming redundant, while a system that is too far out of the range of a user's experience risks being of little interest. Novel systems help attract users, in part, by leveraging their natural curiosity. And, since new experiences are engaging, system novelty can give a boost to the engagement of first-time users; this boost lasts until they become familiar with the system.

As indicated by fig. 3.3, once a user has decided to try the system, the question of how long a session he or she will choose to have is partially determined by outside pressures specific to each individual user. Outside pressures, such as responsibilities, physical needs, or time constraints, increase in urgency as they go unaddressed. The session ends when the outside pressures overcome the user's desire to continue the session. What determines the level of that desire to continue, in the early stages, are factors falling under the purview of design. Among these factors are targeting and novelty. However, there are two other factors which may be more important. The first is *cognitive richness*, which I define as "how well the system captures user attention." The mediated activity afforded by the system should be matched to the attentional bandwidth of early-stage users, meaning these users should find the system challenging. This will help inspire a state of flow. Although there is nothing in the framework to dictate visual or musical aesthetics, this heuristic implies that systems will be more successful if they are rich in sensory information, as these will elicit more immersion. The second factor is *usability*. Although the system's afforded activity should be challenging, participating in that activity should be hassle-free. Any distractions from the afforded activity will remove users from the flow state, cause them to disengage from usage, and reduce the likelihood of future attempts. Bugs and crashes are the most egregious of typical usability problems, but other distractions

70

include unnecessarily difficult operations, unnatural interactions, or confusing interfaces that cause user errors.
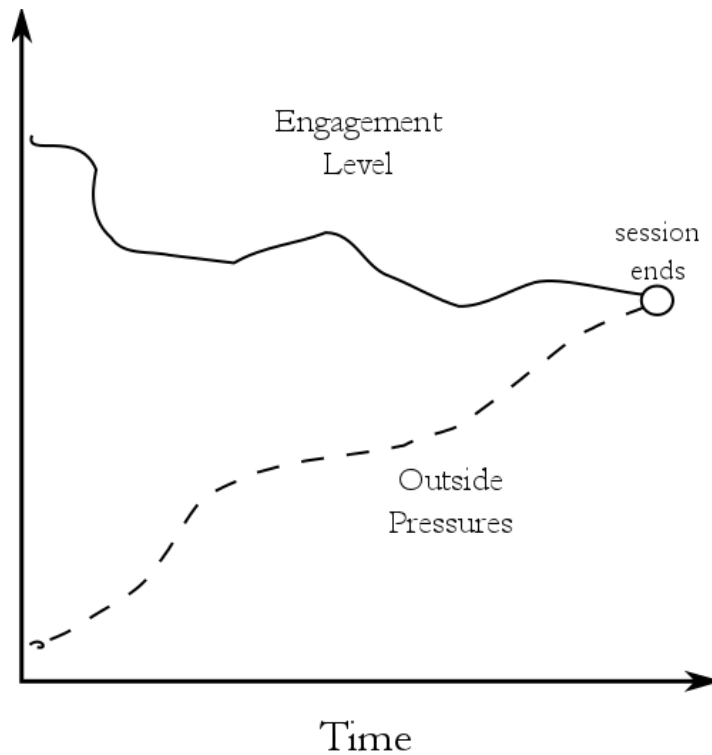
Figure 3.3: Example depiction of a voluntary usage session with regard to when the session ends.

Summing up the discussion of short-term engagement so far results in a set of heuristics as seen in table 3.1. Obviously, there is great individual variance within any desired user base, so no matter how well-designed the system, some users will remain uninterested and unengaged. It is similarly unlikely that the number of unaware or unavailable potential users can, by any means whatsoever, be reduced to zero. There is one method for further reducing the numbers of these groups that has not, thus far, been discussed, which is incentivizing usage with reward. If longer-term volitional usage is part of the system goals, this is not recommended: as discussed in section 2.3, extrinsic motives depress intrinsic motives. Activities that draw longer-term participation, such as hobbies, are done entirely out of enjoyment. This type of engagement is

only observable when people choose, of their own intrinsic volition, to participate. It cannot be verified in the presence of rewards or punishments.

3.2 Longer-term Usage

The factors governing volitional usage in first-time users are not identical to the factors governing volitional usage in continuing users. One way to understand the relationship between the two sets of factors is through a weighted model where, at the beginning of a usage career, the shorter-term factors are heavily weighted and the longer-term factors are lightly weighted—but the weighting balance shifts as time progresses and the user participates in more usage sessions. The short-term factors drive usage less and less, while the longer-term factors drive usage more and more. If a system is not as well-designed with regard to longer-term factors as shorter-term, then its users will tend to quit as time progresses. This model is too simplistic, because some of the longer-term heuristics are loose corollaries of short-term heuristics, so the idea of "switching over" from one heuristic to another is problematic—rather, it is as if a single set of heuristics transforms itself in complex yet continuous ways over time. Still, it is useful for designers to consider these two types of usage separately, so this dissertation provides parsimonious design frameworks for both.

Motive	Heuristic	Description	Impact
Mastery	Incrementality	Gradual open-ended learning curve	Maximizes state of flow over time
	Immediacy	Usable, convenient, low costs	Impacts usage initiation & persistence
Autonomy	Ownership	Users can express & invest themselves	Gives sense that interface is well-suited
	Explorability	Users have operational freedom, complexity	Without explorability, boredom results
Purpose	Demonstrability	Skilled users have something to share	Incentivizes mastery, draws new users
	Cooperation	Users can participate together	Fosters sharing and motivation

Table 3.2: Heuristics for longer-term volitional usage.

Due to the nature of longer-term volitional usage, the provenance of the longer-term heuristics differs from that of the short-term heuristics. The longer-term heuristics are more based on self-determination theory and studies of hobby. For each of the three intrinsic motives in self-determination theory—mastery, autonomy, and purpose—this framework contains two heuristics. A brief summary of each of these heuristics can be found in table 3.2. In upcoming subsections I cover each heuristic along with any logical relationships between the heuristic under discussion and previously-discussed heuristics. I also cover how people might behave when systems are not well-designed with regard to the heuristic in question. As discussed in section 2.2, heuristics imply idealized characteristics. No system can be perfected with regard to any heuristic, and although it is useful to discuss designs in terms of heuristic quantities (e.g. "This system has more immediacy than that one."), this risks being misleading because in most cases there will be no simple comparison. Users vary with regard to the levels of importance placed on each characteristic discussed. Some users may care little for cooperation, while others find cooperative activities engaging, for example. User impressions on the implementation of specific heuristics will vary too. For example, some users may feel that a system provides a sense of ownership, while other users of that system do not. In some design circumstances, practicalities may dictate that certain heuristics must go unaddressed. Some people can, through time and effort, overcome deficiencies with regard to any heuristic, often creating the missing characteristic for themselves in the process. However, the fact that some users might overcome problems with the heuristic qualities is no excuse for disregarding any heuristic thoughtlessly, because the number of users who overcome such deficiencies will probably be smaller than the number of people who would otherwise have engaged in longer-term usage.

The conceptualization of these heuristics is based, in part, on anecdotal evidence obtained from looking at certain hobbies—especially musical hobbies—in the context of self-determination theory. In these hobbies, the intrinsic motives of mastery, autonomy, and purpose emerge from the characteristics underlying these heuristics. To illustrate this, examples

from the field of hobby are given in the discussion for each heuristic. The fact that musical hobbies are popular suggests that this set of characteristics is compelling. Therefore, these characteristics can be used as heuristics to help designers organize their thinking during the design process. In providing this longer-term framework, I posit that considering these heuristics, and trying to incorporate them into systems, will result in systems that sustain volitional usage more effectively than if the heuristics are not taken into account.

Immediacy: low participation costs

In looking at the discussions of convenience and usability in section 3.1, one can make the case that they are each related to the costs of usage. Here 'cost' is defined broadly—the term includes effort costs and time costs as well as the customary financial meaning—but is delimited to things that are required for participation in an activity, but are not directly part of that activity. Therefore, under this definition, doing the work of pedaling a bicycle is not a cost of cycling because it is integral to the activity; but purchasing, maintaining, and locking and unlocking the bicycle are costs of participation. Delays, distractions, inconveniences, and expenses are all forms of participation cost. In this dissertation, the term 'immediacy' is used to denote all costs of participation, including costs related to convenience and usability. A lack of immediacy restricts usage greatly—it is frequently disqualifying to users. However, the quality of immediacy does not impact whether an interface will be interesting or fun, it merely ensures that it will not be frustrating or inaccessible.

As discussed in section 3.1, convenience and usability are important predictors of whether users will choose to try an activity and whether they will enjoy themselves in the short term. Unlike short-term heuristics such as novelty, however, cost-related characteristics impact volitional usage throughout the usage career. Even previously-dedicated participants might quit an activity if that activity suddenly became less convenient, for example. Of course, participation costs are often more limiting in early stages of usage because: a) many costs, such

as purchasing or acquisition costs, are one-time only; and b) with commitment, people can become accustomed to costs over time. Consider swimmers: many swimmers do not own a pool, so they must travel to one. Yet this is not much of an obstacle for dedicated swimmers; for them, traveling to the pool has become an acceptable cost of participation. On the other hand, swimming pool access probably does drive a socioeconomic bias in swimming sports.

Musical instruments provide interesting examples of the relationship between volitional usage and immediacy. Some instruments are practiced as a daily hobby, in part, simply because they are easy to practice in this way. These instruments can be left on a stand in a convenient location, so picking them up and playing a few notes takes little effort. If these instruments required preparation or delays before each practice session, they would not be as popular. Traditional harps have less immediacy than harmonicas because they are not portable and much more expensive. Traditional organs have less immediacy than guitars because guitars are fairly ubiquitous while organs are rare and fixed to specific sites. Software instruments have less immediacy than flutes because of the delays involved with starting up the computer and launching the software.

Since immediacy impacts the likelihood of people taking up an activity and also impacts the likelihood of them persisting within it, designers should note that a relationship will exist between immediacy and the size of the user base. Therefore, if other things are equal, one way of promoting an interface's success is to ensure it has the quality of immediacy. In many cases, this will entail careful selection of the HCI delivery vehicle, as this will impact many factors relating to immediacy, such as: latency and start-up times, ease of setup and use, portability or ubiquitousness, and how expensive an interface is to obtain. For example, immediacy is a common factor among many recent successful interfaces delivered via mobile technology.

Incrementality: gradual learning curve

The motive of mastery is due in part to the pursuit of flow (Csikszentmihalyi, 1991) which is, as discussed in section 2.3, an enjoyable mental state attainable through performing activities that are complex but not overwhelming. Extrapolating from this, it follows that the challenge involved in any longer-term activity inducing flow will likely conform to a specific profile over time: it will start small for beginners, then increase gradually as skill is attained (Figure 3.4). If challenge increases too quickly, participants may become overwhelmed and quit, and if it increases too slowly, participants may become bored and lose interest. Therefore, the term 'incrementality' denotes the gradualness of a system's challenge profile. The way incrementality is manifested varies from activity to activity. Video games, for example, manage challenge through levels and scoring. Instrumentalists, on the other hand, manage their own challenge levels: there is such a diversity of music available to learn that musicians at all skill levels typically have an abundance of appropriately difficult music.

Incrementality is related to the short-term heuristic of cognitive richness. As discussed in section 3.1, systems are cognitively rich if they monopolize the attentional bandwidth of users—thereby inducing immersion or flow—through complex feedback or challenging activity. This characteristic remains important throughout a usage career, but can only be maintained through incrementality. Flow and immersion depend on cognitive skills held by the user, and these skills will improve with experience in the activity. When users get good enough at performing a mediated activity, they stop being challenged and cease experiencing flow. Similarly, when users get good enough at processing multimodal information, they cease being immersed. Certain forms of media, such as movies and television, are capable of maintaining a state of immersion for longer periods of time because they provide an influx of new information, often in the form of narrative. In activity-based forms of media, however, sustaining a state of flow requires a slow but constant ramping-up of the challenge involved.

(a) Ideal profile for longer-term engagement.

(b) Users may grow bored in the short term.

(c) Users may be overwhelmed at the outset.

(d) Users engaged, then overwhelmed.
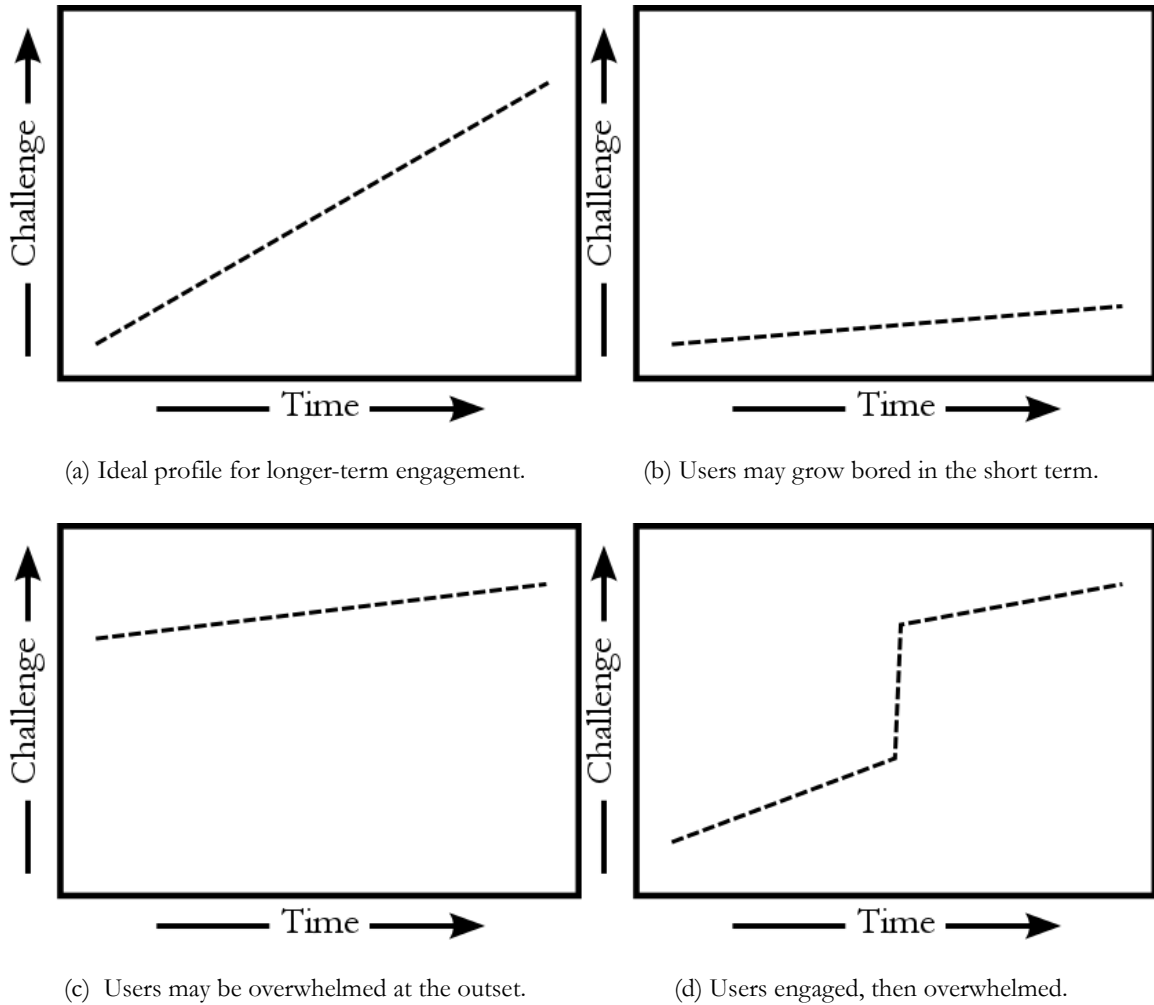
Figure 3.4: Incrementality profiles.

Just as incrementality implies cognitive richness should be extended throughout a usage career, it should be noted that gradual learning curves can and should be compressed into single usage sessions for more cognitive richness in shorter-term contexts. Examples of shorter-term incrementality include system tutorials. If well-designed, they allow users to progress quickly to learning things which are personally challenging, and once they master that difficulty level they can move on to a higher one. If gradual learning curves exist within single sessions, users will be more willing to continue each usage session for longer periods of time.

There are differences in incrementality between instruments, and these differences illustrate relationships between incrementality, the size of a user base, and the dedication of a user base. For example, it takes practice even to play one's first notes on an oboe. There are fewer players of this type of instrument than instruments that are easy to play notes on, such as piano. However, dedication is probably higher among oboe players as they have invested significantly more effort from the outset. Other instruments, such as diatonic harmonica, are easy to take up, but have large challenge jumps corresponding with times when advanced techniques such as note bending are learned. These jumps in challenge are large enough that some people set the instrument aside instead of learning the advanced techniques. Many musical instruments have known plateaus in their learning curves, yet these instruments are still played. Therefore, even if a system does not have an idealized heuristic of incrementality, some users (provided there are other characteristics of the system engaging these users) will overcome this through sheer persistence.

Ownership: self-expression and self-investment

Ownership occurs when users associate systems with unique forms of self-expression, make a significant investment of self in the form of time or effort, or (in many cases) both. One way to transfer the concept of ownership to HCI design is to integrate options and end-user configurability into the system. Another way, used frequently in video games, consists of rewarding accomplishment of system goals with access to new customizations, interactions, or advanced features. A less obvious tactic for transferring the concept of ownership to HCI is advocated by Sengers and Gaver (2006). In their formulation, interactive media should be designed so that it can be appropriated; i.e., so that users can define its meaning for themselves. In order for appropriation to take place, there must be something ambiguous or abstract about the media, leaving room for personal interpretation so that most users will not interpret the same system in the same way. Ownership's closest corollary among the short-term heuristics is

targeting, because both ownership and targeting imply overlap between user interests and the system's afforded activity. With targeting, however, this overlap exists because the system was designed according to user interests, whereas with ownership this overlap exists because the user has transformed the system, and transformed him-or-her self, over time. The user and system have grown together according to the user's wishes. To explain this difference another way, system targeting is based on design features, which are pre-determined, whereas system ownership is based on design affordances, which can be exploited by users in various ways.

In musical instruments, mastery and autonomy are related. The diversity of music affords its incrementality and complexity, and also affords a completely individualized path of learning for instrumentalists. Ergo, each musician can consider their playing style to be unique, best suited for them, and therefore "owned" by them. Renowned masters of music provide case studies on stylistic ownership: for example, Franz Liszt and Art Tatum are both considered absolute masters of the piano, yet their styles and techniques were highly dissimilar. The sense of ownership is one factor allowing musicians to consider their music to be a form of self-expression. Playing styles are developed with great effort over the course of long periods of time, so ownership in instrumental performance also represents investment that may deter musicians from quitting. We are reluctant to set aside hard-won skills. For this reason, ownership has an inverse relationship with the heuristics of immediacy and incrementality, because the sense of ownership is acquired, in part, through paying participation costs and overcoming difficult challenges. If participation costs are low and challenges are never too difficult, then activity participation may not develop one's sense of ownership. On the other hand, more people may be inclined to participate in this easier activity—but this could make it harder for individuals to feel as though they perform the activity in their own unique way.

To expand on the relationship between ownership and incrementality, as people progress along the learning curve of an activity, the more of an affinity they feel for that activity. If they participate long enough, a sense of buy-in will start to accrue. Their skills will develop, and

a sense of ownership will result. This can take place whether the activity was well-targeted to the user, in the early stages, or not: the original interests and predilections of longer-term participants will not matter as much, because these interests and predilections will expand to include the new activity. This phenomenon is one rationale of parents who make their children take music lessons. The children may not initially wish to take the lessons, and may not feel engaged during their first practice sessions. Sometimes, however, the children will grow to appreciate both the lessons and the practice sessions as they spend more time developing skill and expressive capability within the activity of music-making. Sometimes not.

Seeing that a sense of ownership usually takes time and effort to develop, this heuristic is more observable in longer-term usage contexts than in shorter-term usage contexts. Even in cases where people perform customizing or personalizing activities in shorter-term contexts (for example, when video-gamers spend their first game-play sessions customizing an avatar to represent them in-game), longer-term usage is often the driving force: it seems unlikely that video gamers would spend much time customizing avatars unless there were an intention to use those avatars for significant lengths of time. Conversely, even when interfaces are not explicitly designed for ownership, given time users may find ways to create the sense of ownership for themselves. For example, there is usually little functional difference between instruments of a type. One guitar plays similarly to other guitars; ownership is not built into the guitar on any level except the cosmetic. Yet guitar players feel a great sense of ownership with regard to the activity of playing guitar, because they have invested time and effort into learning the instrument, and because they have specialized in playing music they enjoy.

Explorability: operational freedom and complexity

In many activities, the term 'skill' is equivalent to "capability of accomplishing complex things." This is the case in instrument playing, where there is such potential complexity that no individual can fully master any non-trivial instrument; some facet of playing the instrument could

always be improved. This impacts volitional usage because it means that the upward-trending challenge profile associated with incrementality extends into infinity, so participants can theoretically remain engaged with usage forever. When activities do not seem to afford much complexity, they are often perceived as trivial and are usually not the source of longer-term engagement. For example, kazoos seem toy-like and simple to play, therefore they are not widely played as a hobby. The complexity of an activity can be thought of as the ceiling or the end-point of the incrementality profile.

Despite its counterintuitive nature, the idea that interactive complexity is a positive design trait is not new to the field of HCI: for example, Norman (2010) provides an exhaustive argument in favor of design complexity. Systems which are needlessly and egregiously complex will be less usable, so complexity and usability need to be balanced, but the two characteristics can and often do coexist. Beginners should find it simple to identify and use the core operations of a system, and extraneous or confusing interface elements should be avoided, but the potential for complex interactions should exist for users with sufficient skill levels. The afforded activity of a system should not demand endless repetition from users, and the system feedback should not be too repetitive.

At one point in the history of this heuristic framework, complexity and operational freedom were considered to be individual heuristics (cf. Wallis et al., 2013). However, in looking at musical instruments, it eventually became apparent that these concepts are not orthogonal. As interfaces, instruments are highly simplistic. Most have only a few interface elements in the form of strings to pluck or membranes to strike. Yet the potential complexity of most instruments is effectively infinite, and this complexity emerges from the fact that instruments afford a great deal of operational freedom. When an interface affords numerous ways to interact which can be applied in any order or at any time, that interface has operational freedom. Some users may not use their operational freedom to do complex things, but potential complexity exists nonetheless, because users are capable of combining or sequencing interactions in innovative

ways to create complex outcomes. Therefore, when operational freedom is high, complexity is also high, so integrating operational freedom into systems is one way to neatly solve the complexity versus usability problem discussed in the last paragraph. Integrating operational freedom may also help a sense of ownership to emerge, because operational freedom affords users the opportunity to develop personal usage styles. One reason complexity and operational freedom are so tightly coupled is that they each work to encourage user engagement through leveraging the user's natural desire to explore; to see what happens next; to see what the difference is between *this* complex interaction and *that* one. In this they are like the short-term heuristic of novelty, which similarly leverages user curiosity.

Like the heuristic of immediacy, explorability often goes unnoticed when well-designed. It is only when the limits have been perceived—when there is nothing left to explore within a system—that this heuristic becomes visible to users, and at this point usage may drop precipitously. For example, in the field of video gaming, once players have played through a game, they will probably play it less. However, we see by the existence of certain hobbies that limits to explorability can be overcome, sometimes through subverting the original intent of the activity. Trick bicycle riding exists because some people went beyond the perceived limits of freedom and complexity afforded by regular bicycle usage, by constructing ever-greater challenges for themselves with regard to jumping and balancing. These originators disregarded the custom, dictated by the form and affordances of the bicycle, of sitting on the seat with the feet on the pedals. By doing these things, they also subverted the paradigm of cycling for transportation.

In some design circumstances, specific user interactions are desired, either because of the application or because of other factors such as constraints in the sensing technology. For example, movement rehabilitation systems often need to encourage beneficial movement habits while discouraging poor movement habits (e.g. Wallis et al., 2007). Designing systems that have operational freedom can be challenging in these circumstances, but musical instruments provide an applicable model. As noted in Jordà (2002), instruments have affordances, and

these affordances lead to stylistic similarities among the players of any given instrument. Expert musicians are sometimes capable of going beyond the natural affordances of an instrument, but in most cases playing techniques will converge toward efficiency and ergonomics. Transferring this concept to HCI development, when specific user interactions are needed, designers can carefully sculpt interface affordances so that users will gravitate to desired interactions without perceiving restricted operational freedom. This insight is not new to the field of HCI: many theorists have expounded on the topic of designing affordances (e.g. Gaver, 1991; Norman, 2002).

Certain media experts have been advocating against the engineering of user experience in interactive media, and for greater autonomy. Sengers et al. (2008) defines a new media-affected mental state, called enchantment, which has similarities with engagement and immersion. When users are enchanted with a system, they are engaged without perceiving any logical underlying structure. They do not know how the system works, and probably will not ever know, yet are captured by the mystery of the experience. In an ideal system, even the system designer may not know with certainty how the system works. The ambiguity of such a system should render it incapable of dictating to its users, so enchanted users become engaged with freely exploring and trying to understand a rich and complex environment. Although designing a system without understanding its operation may seem like a task with insurmountable logical inconsistencies, the intricacies between emergent human behavior and system form make it possible to design unpredictable systems which provide an air of mystery and surprise. Computational modeling algorithms such as rule-based models can also be integrated into systems to add more unpredictability.

Demonstrability: ability to exhibit skill

People often take up hobbies in order to attain skill and then demonstrate that skill to others. This is true for hobbies such as juggling, and it is also true for hobbies based in the arts. For

example, people often learn instruments for this reason. New musicians may choose to take up an instrument because they have heard impressively performed music on that instrument (Manturzewska, 1990). Similarly, if interfaces are designed such that users produce something that can be displayed, performed, or shared in some way, this will encourage users to attain greater levels of skill, and these users may impress and attract more new users. Of all the heuristics, demonstrability is the most difficult to defend as an intrinsic motive, because it seems related to an extrinsic motive based on the approval of others—in a way, audience approval is the payoff for attaining expertise. However, demonstrability is also related to the desire to express oneself, which is more of an intrinsically motivated desire.

Demonstrability facilitates both shorter and longer-term usage, but its relationship with longer-term usage is strong because those who have participated in an activity for longer are more capable of demonstrating skill. The more skill people have in an activity, the more likely it is that they will want to demonstrate it. This is a double-edged sword, however, because audiences can sometimes reduce the desire to participate in an activity. For example, sometimes unskilled users may experience performance anxiety, and sometimes skilled users may not wish to be perceived as show-offs. There is great variance between individuals with regard to the inclination toward demonstrability, and many activity participants are perfectly happy participating in solitude. For this reason, it is advisable that systems be designed so they can be used both privately and publicly. That way, people can practice alone if they choose, but need not always do so. This is essentially an argument in favor of portability.

Of course, the presence or absence of an audience is often more of an environmental externality than a design feature, and this means it is possible to make any system more demonstrable merely by bringing in people to watch while the system is being used. This may not be practical over multiple usage sessions, however, so this is a tactic for impacting shorter-term volitional usage rather than longer-term. However, this can be an effective means of reducing problems related to awkwardness or performance anxiety. That is, people are more likely to

engage in usage sessions if given an audience that is: a) involved cooperatively with system use, perhaps by giving tips or suggestions to the user; b) of a similar skill level to the user; and c) supportive, friendly, and ideally, trusted by the user.

Cooperation: ability to participate with others

Music making can be done in solo or ensemble settings. The option to play in ensembles contributes to the intrinsic motive of purpose, as musicians are often motivated to practice by the prospect of jam sessions, drum circles, duets, and so forth. These represent social opportunities, allowing players to spend time with peers and make new friends. As noted in Swift et al. (2013), musicians often describe a shared feeling of euphoria, immersion, and engagement when playing or improvising music well together. Cooperation also allows musicians to teach one another, inspire one another, and motivate one another. If interfaces are designed to be used in group settings, and efforts are made to increase community among users (for example, through online forums and wikis) this will help increase overall engagement within the user base. It will also help attract new users and speed the attainment of skill in the user community as a non-competitive environment of knowledge sharing and group discovery develops. According to the theories on student persistence outlined in Tinto (1987), social integration will also reduce an individual's likelihood of quitting.

It is important to note that this heuristic is defined broadly, and includes circumstances beyond traditional teamwork-based activities (such as ensemble music-making). For example, the heuristic of cooperation includes competition. A pair of chess players, despite their mutual desire to defeat each other, are cooperating in many ways. They have agreed upon an activity to participate in, a set of rules to abide by, and presumably they are each interested in ensuring everyone involved has a good time and acquires skill at playing chess. I would argue that cooperation takes place even in single-person activities in the circumstance that participants communicate to share knowledge or get together to work on separate projects (as in knitting

86

parties, for example). The simple presence of co-participants adds a motivational element to the activity. This shows that the characteristic of cooperation is often purely environmental. Activities can become cooperative not only through inherent features but also through the fact that interested participants form communities. For example, many video games have no social elements, and the players of these games often play in solitude. However, should they choose, these gamers could join tournaments, chat with each other on internet forums, or create video walk-throughs to share with one another online. There are many successful interfaces that have no design features explicitly addressing cooperative use. Activities which are cooperative are more engaging for many users, but if cooperation is not built into a design this may go unnoticed. If systems are not designed for cooperative use but manage to draw a number of passionate users anyway, internet forums and physical meet-ups may be organized, and a cooperative characteristic will emerge on its own.

The presence or absence of a group of activity participants is often more of an organizational externality than a design feature. Consequently, it is possible to make almost any system more cooperative merely by getting people together to use it. Although this tactic is viable in longer-term situations, it is easier to employ in shorter-term situations. To explain why, one need only look at musical instruments. Musical instruments are cooperative only when musicians get together to play. Sometimes ensembles play together regularly over the course of years, but inconvenience and individual pressures will tend to cause ensemble numbers to dwindle over time. Without an influx of new members, the ensemble will eventually go defunct. The presence of group members willing to form organizational structures and take on organizational roles can greatly lengthen the life-span of a cooperative effort.

3.3 Application

In this chapter, heuristics have been presented to facilitate the design of systems that draw volitional usage in the short term and in the longer term. Tables 3.1 and 3.2 summarize these heuristics. I attempted to create a balance so that heuristics were defined broadly enough to minimize intra-framework redundancy, but were also defined narrowly enough to give designers useful, simple, self-encapsulated concepts to work with. For example, I might have combined incrementality and demonstrability into a single more broadly-defined heuristic related to skills held by the user. I chose not to, because these two heuristics impact different intrinsic motives: incrementality relates to skill attainment (mastery) while demonstrability relates to sharing with others (purpose). With a single combined heuristic, designers might overlook factors that are important in the context of these underlying intrinsic motives. In another example, immediacy already encompasses both convenience and usability, so these heuristics could be combined. Convenience and usability address different problems causing early-stage user forfeit, however, and it may be important for designers to consider these problems separately; therefore I decided to treat convenience and usability as distinct heuristics in the context of short-term usage. Acknowledging that conceptual interrelationships exist between the various heuristics, table 3.3 provides a loose reconciliation between short-term and longer-term sets of heuristics that attempts to minimize conceptual redundancy within the overall framework. This results in a more manageable number of heuristics, but sacrifices the conceptual separation between short-term and longer-term types of volitional usage.

Whenever HCI developers design systems that could benefit from longer-term engagement, heuristics from this chapter can be used as catalysts for thought. Developers should ask themselves questions like: "Is this system demonstrable?" or "Would different sensing technology make this system more usable, and therefore more immediate?" Framework heuristics can be considered at any stage of design, to include the preliminary idea generation phase and

Desirable Characteristic	Implications
Catches potential user's eye	implies novelty, aesthetics, promotion, targeting
Has low participation costs	implies convenience, usability, and immediacy
Uses attentional bandwidth	implies challenge, rich feedback
Gradually builds a sharable skill	implies incrementality and demonstrability
Fosters self-expression, self-investment	implies targeting, affordance of ownership
Offers endless exploration	implies novelty, operational freedom, complexity
Affords social usage	implies cooperation and demonstrability

Table 3.3: Loose reconciliation between short-term and longer-term design heuristics.

the concluding product analysis phase. Maximal utility might be drawn from the framework if it is applied at the outset of the design process, when the designer has done nothing except identify a human-centered problem to be addressed. This will help avoid premising the interface on some application or afforded activity that is not conducive to longer-term engagement. Designers using this framework will tend to create interfaces that have creative elements, game-like elements, or elements of knowledge or skill building. Not coincidentally, one or more of these elements are found in essentially all hobbies. However, if for some reason the interface's application disallows all these elements, the framework may prove to be of little assistance.

When using the heuristics to inform the preliminary idea of an interface, it is useful for designers to address the human-centered problem from the perspective of creating engaging activities, rather than engaging interfaces. In other words, the interface being developed should be thought of as a portal or facilitator to an engaging activity. This is helpful in part because there are numerous existing hobbies and activities that people find engaging over long time periods. These can be mined for ideas: if a compelling activity already exists that can be mediated, designers may be able to create a modified version of that activity addressing the human-centered problem. For example, knowing that many people find word puzzles compelling, a designer could address the human-centered problem of foreign language learning by making interactive word puzzles that help users teach themselves foreign languages. Although

this framework could be used to design systems that address no human-centered problems—systems that exist only for user entertainment—I recommend that practice-oriented systems designers do take on human-centered problems, for two reasons. First, addressing human-centered problems will constrain the design possibilities, and this often helps designers avoid tyranny-of-choice issues. Second, many engaging self-expressive hobbies already exist; if users want a hobby for entertainment purposes only, they could simply take up one of these existing hobbies. Practice-oriented interfaces are only innovative and fruitful when they foster transformative practices.

Chapter 4

THE FRAMEWORK AS ANALYTICAL TOOL

In order for the design framework laid out in the previous chapter to be successful, it must assist interaction designers in two ways. First, it must provide an intellectual basis for analyzing systems. Second, it must provide a set of guidelines when creating new systems. This chapter is designed to show that my framework satisfies the first requirement. A series of interactive systems will be discussed here. As it happens, I had a significant role in the creation of all of these systems, and they were all designed to elicit usage and to be engaging, and working on them eventually led to the creation of my design framework. My involvement with all of them, however, was finished before I ever began formally looking at the topic of volitional usage. This makes it impossible to perform any formal evaluation of these systems with regard to the design framework, but it also makes them good candidates for intellectual analysis using the framework, because a) I am deeply familiar with the goals and operation of these systems, and b) I can recall and describe my impressions of how these systems were received by users.

Of course, such impressions provide no real evidence that my framework has value for designers. The only things that could provide such evidence are the analyses themselves. In qualitative research methodology, there is a widely-known method of research called *grounded theory*, in which qualitative data such as interview transcripts is used to generate new theories (Glaser and Strauss, 2009). Grounded theories are judged according to four criteria: relevance (how useful the theory is outside the realm of the purely academic), fit (how well the theory fits the data), modifiability (whether the theory can be modified in the case of new relevant data), and workability (how well the theory findings can be used for explanation or prediction across variable circumstances). These criteria also seem applicable to my design framework. The rel-

evance of my framework is discussed at length in chapter 1. Since there is little data (other than my recollections and published papers) on the systems in this chapter, they provide little help in validating theory fit or modifiability—these criteria will be explored in chapter 5. However, analyzing these systems will go a long way toward validating that my design framework is workable.

Each of the systems discussed was designed to address specific functions. These functions will be discussed. However, there were also goals which were common to all. These goals include:

- **Drawing usage and engagement:** All discussed systems took widely different approaches to drawing volitional usage and inspiring engagement in users. In my analyses, I will be discussing the details of the different approaches and any successful or unsuccessful aspects. Special attention will be devoted to the relationship between aspects of the design framework and the way users responded to the system.

- **Aesthetically compelling:** All discussed systems have artistic value, and each system contains a musical element. In my analyses, I will be discussing the overall aesthetics of systems, to include system musicality (e.g. musical mood, interestingness, and information transference). Special attention will be devoted to the aesthetics of each system's afforded activity. For instance, when possible, afforded activities will be compared to similar activities which previously existed.

- **Informing the theory and praxis of experiential systems design:** One of my personal goals for working on each of these projects was to develop practical skills and insight on the topic of experiential systems design. In my analyses, I will be discussing any insights about the process of experiential systems design, including suggested practices, gleaned from working on these projects. Special attention will be devoted to instances where one of these systems impacted the creation of the design framework.

4.1 Adaptive Mixed Reality Rehabilitation

This section describes the first research project I worked on as part of my graduate study. This project was a large-scale team effort, where the team included Dr. Thanassis Rikakis, Prof. Todd Ingalls, Prof. Loren Olsen, Dr. Yinpeng Chen, Weiwei Xu, Dr. Hari Sundaram, Dr. He Huang, Dr. Jiping He, Dr. Margaret Duff, and many others. This research project was in existence for several years before I joined it. During the four years I was involved, we researched and iteratively developed a task-based experiential media biofeedback system. This system, as one of its elements, incorporates musical feedback as a way to maintain patient interest and impart movement information to the patient. My previous background was in music and music programming, hence my role in this project was as a designer of the musical sonification engine. With guidance from Prof. Todd Ingalls and Dr. Thanassis Rikakis, I designed algorithms and programs that play musical feedback in response to patient movements. I describe these algorithms and programs in Wallis et al. (2007), and much of this section incorporates language from that publication.

AMRR is a movement-based system designed around the idea that, with enough accurate movement information in the form of real-time aesthetic media, rehabilitating listeners can remap mental pathways to limb control, thus speeding and improving the process of stroke rehabilitation. AMRR focuses mainly on patient efficiency, fine control, and speed during a right-arm reaching task. Patients who have suffered a stroke and lost some limb functionality may not easily regain the cognitive processes by which they used to control their own movement; but, by using graphics and sound to provide accurate movement feedback they can remap their movement control through other cognitive processes which are exercised during the physical therapy. Simultaneously, the system is meant to provide a greater level of engagement than traditional physical therapy.

93

This system uses animated visuals and sonification to provide information and incentive for better movement. It is designed for patients who have right-arm impairment due to stroke, who have also been in recovery for six months or more. Research exists indicating that the use of interactive feedback in conjunction with movement therapy has a significant impact on the functional recovery of stroke patients with sensorimotor deficits (Dombovy, 2004; He and Jiping, 2006; Liepert et al., 2000). However, there are some significant limitations affecting the outcome and validation of these studies: a) Most of the systems relied on unimodal feedback or simple combinations of modes; b) feedback was informational and superficial rather than experiential and structural (for example, a system might alert the patient that their movement speed was too fast but would not provide feedback that slowed the patients movement in an intuitive manner); c) these systems were not adaptive (e.g. it was not possible to change the therapy based on individual abilities and rate of progress), and d) the test populations were small.

Before I joined the project, the team did a study where an interactive multimodal environment was designed for repetitive reaching and grasping retraining (Chen et al., 2006). In this environment an animated arm tracks the movements of the patient's real arm, but the animated arm could move through different visual scenes. The patient could grab objects in these scenes using the animated arm. There were two primary environments. The first was a virtual table with objects like teapots and cups, and the second was a fish-tank with fish that the patient could reach out and grab. Cues about spatial accuracy were shown visually through a semi-transparent cone and line indicating how far off from the correct trajectory the hand was. Audio feedback was provided by the progression through a chord sequence, with register changes mapped to opening of elbow. This helped indicate smoothness of movement and provided incentive for further opening of the elbow to extend reach. The shoulder of the tracked arm was also monitored, and if it moved beyond a pre-determined threshold (as in the case of trunk-contorting compensatory movements) a dissonant collection of notes faded in,

played in a high register by wind instruments. This mapping provided incentive to the user to contain the shoulder movement so the main musical phrase could be heard.

Five hemiparesis patients secondary to stroke were tested using the designed biofeedback system. The results showed that patients could perceive assigned biofeedback parameters. The visual augmented feedback improved the spatial consistency of the endpoint position during reaching. The auditory augmented feedback contributed improvement of the smoothness of endpoint trajectory, and the spatiotemporal consistency of reaching performance. After three to five training sessions, patients indicated faster, smoother, and more applied joint range of motion while reaching. The results were encouraging (Chen et al., 2006), but after looking at ways to improve the system, the team decided that more parameters needed to be looked at, and the most accurate measures for these parameters needed to be identified. It was also decided that the system required a more sophisticated musical sonification engine; one that could effectively communicate multiple measures at the same time. This is how I came to join the project.

4.1.1   Description

Sensing in AMRR is based on infrared motion capture. We use a marker-based Motion Analysis Corporation[1] motion capture system. This sensing system requires a calibration before every usage session, but accurately tracks the three-dimensional location and movement of any number of reflective markers. When people wear these markers, the locations can be used to construct virtual skeletal models that move in synchrony with them. In this case, we require a skeletal model of each patient's right arm and shoulder; therefore, before each usage session reflective markers are adhesively fixed to key points on the wrist, elbow, shoulder, and back on the patient's right-hand side. Then the patient is seated at a table in the center of a dedicated AMRR space that is surrounded by six motion-capture cameras (see fig. 4.1).

---

[1]see http://www.motionanalysis.com/

Figure 4.1: Side view of team member using AMRR.

The skeletal and location data from the sensing component is transmitted to a motion analysis engine. This engine (as described in Chen et al., 2006) calculates real-time movement features from the marker data and broadcasts this analysis over multicast UDP at a rate of 100 frames of data per second. A graphical feedback engine and a musical feedback engine each separately receive this motion analysis. In response, they each provide feedback to the patient: the graphical engine sends a video signal to a large flatscreen monitor, and the musical engine plays audio through a pair of powered speakers. The placement of the monitor and speakers can be seen in fig. 4.2. The motion analysis is also received by an annotation, storage, and retrieval system (described in Xu et al., 2006) that provides a web based interface for annotation and display of data. The annotation, storage, and retrieval system also saves the motion capture data stream, parameter settings for the visual and audio engines, and recordings of audio and video produced by the system. Figure 4.3 shows the composition of AMRR with regard to all
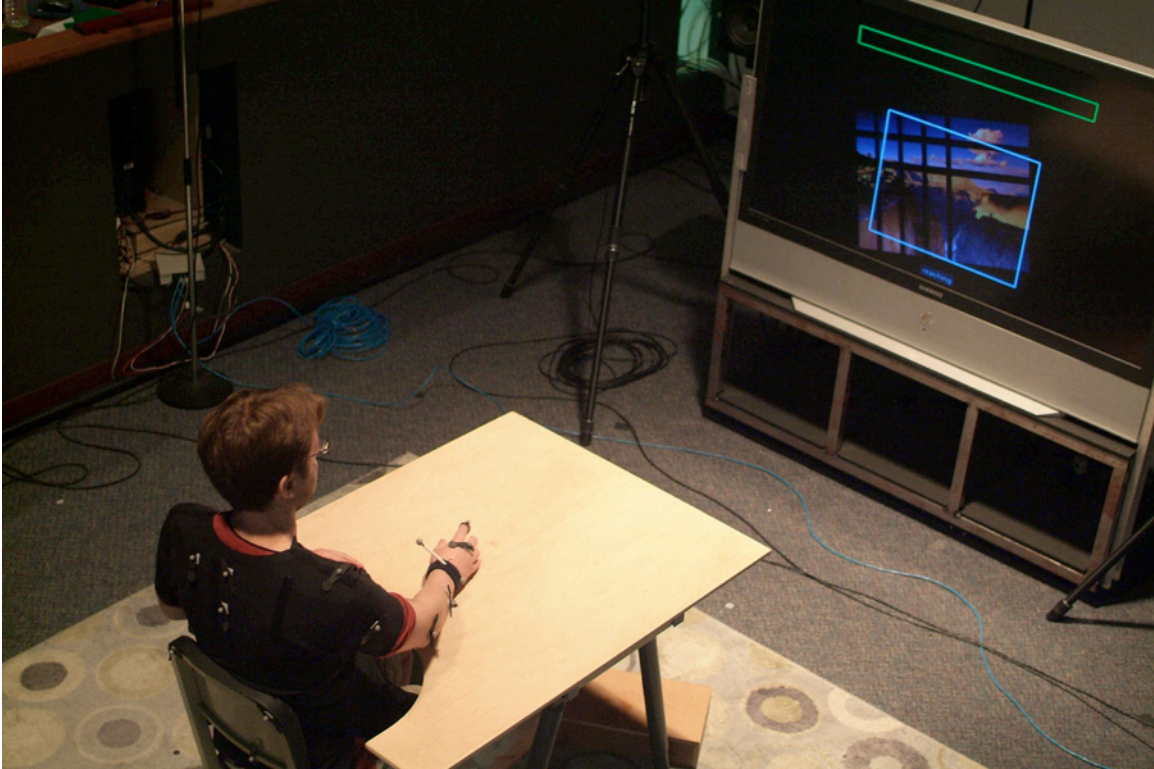
Figure 4.2: View of a team member using AMRR.

of its main components, as an experiential system (i.e. contrast this with fig. 2.2; components of AMRR are similar to components of experiential systems).

The motion analysis engine works, in part, by dividing the effective reaching space into zones (see fig. 4.4), and dividing the user's reaching movement into modes. The exact thresholds used to make these divisions is determined on a per-patient basis by a reaching calibration and by an attending physical therapist. The reaching calibration is performed using a physical object, shaped like a cone, with a reflective marker on top making the cone location trackable by our system. Before each usage session, the therapist places this cone near the patient and asks the patient to grab it. Then the therapist moves the cone a little further away and asks the patient to grab it again. This is continued until a depth-of-reach is found where the patient cannot easily grab the cone without using compensatory movement strategies such as leaning forward. Once this location has been found, it is used to determine the spatial
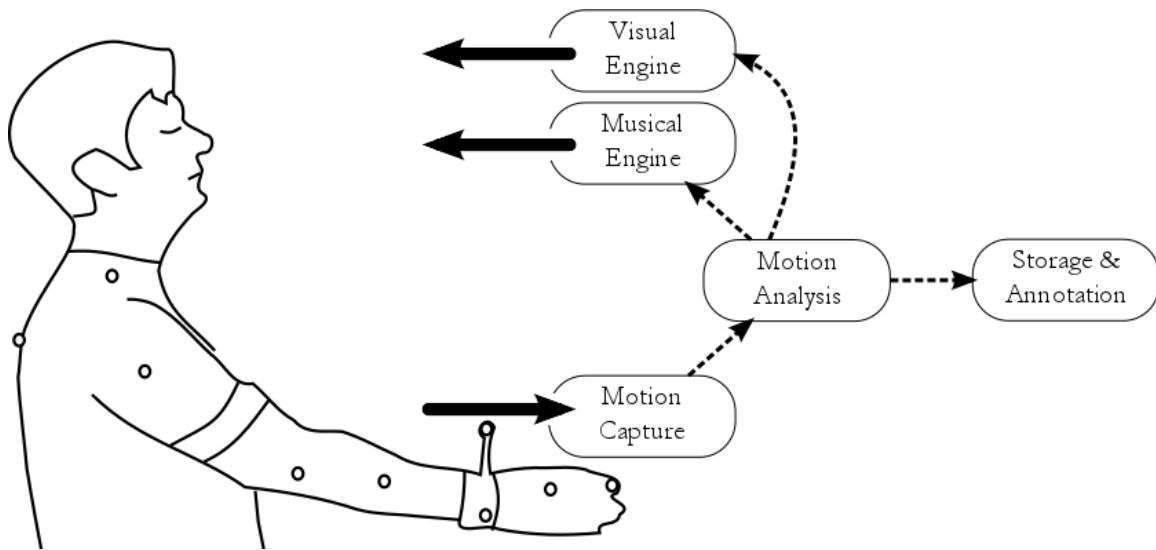
Figure 4.3: Flow of information through components of AMRR.

thresholds defining the zones of reaching space. This ensures the difficulty of the reaching task does not exceed the abilities of the individual patient. The zones are:

- Resting zone: An area on the table near the patient where his or her right arm tends to rest naturally when not reaching.

- Grasping zone: A target area that the patient must attempt to reach.

- Hull: A three-dimensional pathway between resting zone and grasping zone.

These zones allow the motion analysis engine to divide the patient's reach into modes of movement. Usage sessions consist of reaches and sets of reaches. During each reach, patients are tasked with moving their hand from the resting zone to the grasping zone without leaving the hull (the hull expresses bounds for correct arm supination as well). Then, the hand should be held in the grasping zone, with the correct arm supination, for half a second before the arm is brought back to the rest position. If the patient does this, their movement will cycle through the modes listed below. On occasions when the patient is unable to accomplish some part of this task, the reaching mode will not transition. In this case, the modes can be reset by a team member, thereby ending the reach. After each reach, the music and visuals will stop playing for a few seconds during which time the patient is allowed to rest. When the feedback returns,
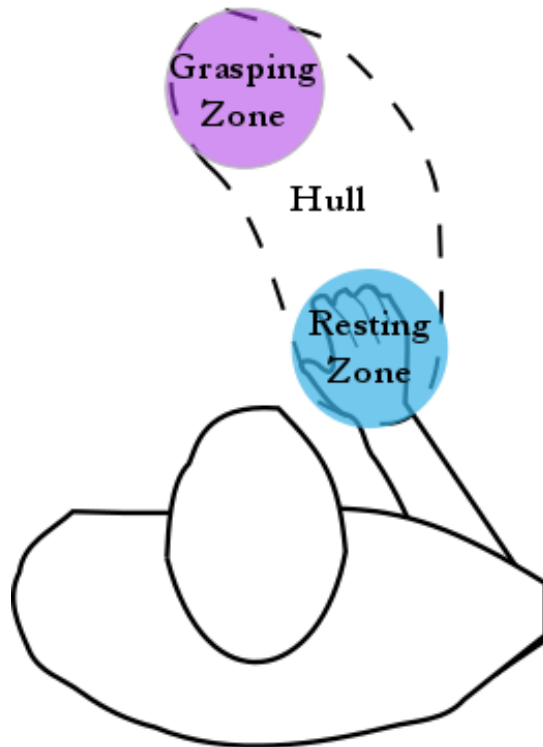
Figure 4.4: Overhead diagram of spatial reaching zones.

the next reach is about to begin. After each set of ten reaches, patients are given a few minutes without feedback in order to rest while discussing with the physical therapist.

- Resting mode: The patient's hand lies static in the resting zone before or after a reach.

- Reaching mode: The patient's arm is in transit from the resting zone to the grasping zone.

- Grasping mode: The patient's hand has reached the grasping zone and stayed there for 0.5 seconds.

- Returning mode: The patient's arm is in transit from the grasping zone back to the resting zone.

In each data frame, the motion analysis broadcasts this mode of reach along with other calculated features such as reaching depth, reaching velocity, wrist supination, openness of elbow and shoulder joints, and many others. All feature calculations are normalized according to the patient's capabilities as ascertained during calibration. For example, the reaching depth is not transmitted in standard units of distance, it is transmitted as a zero-to-one value that is

mapped to the hand location along the length of the hull, so that zero indicates the hand is in the reaching zone and one indicates it is in the grasping zone. Similarly, the reaching velocity is transmitted as a zero-to-one value calculated according to each patient's high speed, so that zero means the patient is not moving the hand, and one means the hand is going fast for this patient.

Here I will discuss how the graphics engine responds to this data. This engine was created using Dash[2], a custom software for animation written by team member Loren Olsen. The visuals in AMRR are designed to give the user a large amount of information, especially with regard to spatial accuracy during reaching. The specific movement features being looked at from the motion analysis are reaching progression and trajectory. There are two visual environments. The first, named the 'explicit environment,' is used to introduce patients to the system. The explicit environment depicts a 3-dimensional virtual room with a table (see fig. 4.5). There is a disembodied right arm hovering over the chair in the same place that the patient's right arm exists in relation to the physical table. The virtual arm follows all movements of the physical arm. A virtual cup exists in the center of the virtual table. This cup corresponds to the grasping zone. The task assigned to the patient is to reach for the cup. Once the patient has touched the virtual cup, and the wrist has supinated in simulation of physical grasping, the cup disappears and the patient may return to the resting position. In this introductory environment, the audio feedback engine is not used.

After patients become familiar with the system, they are switched to the second environment, which we call the 'abstract environment' (see fig. 4.6). In each reach of the abstract visual environment, the patient is shown a picture centered on the screen. This picture changes upon each reach, where the pool of selectable images is configurable per-patient: some patients see images of fine art, others see frames from animated movies, and others see family photographs that they supply themselves. A few seconds after the reaching trial begins, the

---

[2]see http://ame.asu.edu/faculty/olson/

Figure 4.5: Explicit visual environment

picture explodes into particles, revealing a target (represented by a virtual cup, placed in the grasping zone) and a rectangular frame underneath the particle field. The particles move with the patient's hand, and this gives feedback on the progression of the reach, the trajectory of the reach, and the supination of the reach. To give feedback on the progression of reach, the wrist location in Z (i.e. on the length of the hull) is mapped to the spread of the particles. The further the patient's hand is from the target, the more spread out the particles are, and the particle field contracts if the hand is moved forward. If the patient reaches forward until the target is touched (i.e. the grasping zone is reached), the particle field re-forms into the original image. However, if the patient overshoots the target, this image will keep shrinking, leaving empty space between the picture and the rectangular frame—this is meant to tell the patient to pull back. To give feedback on the reaching trajectory, if the wrist leaves the hull in X or Y the

particle field will stretch. The stretch is mapped to wrist location, so if the trajectory leaves the hull, the particle field stretches in the direction of error, and the depth of the stretch indicates the amplitude of error. To give feedback on reaching supination, the particle field rotates, and this rotation is mapped to wrist orientation. In the physical world, a person's hand must roll to a vertical orientation to grasp an object such as a cup, and usually this rolling motion will take place simultaneously with the forward progression of the reach. Therefore, in this system, users must also roll their arm during the reach. The rotation of the arm is mapped to the rotation of the image; if arm rotation is normal, the image will remain oriented as though it were hanging correctly on a wall. If the arm is rotated abnormally at any stage of reaching, however, the image will spin in the direction of the error. This mapping gives the information needed for patients to adapt their arm rotation in order to perform more efficient reaches.

(a) trajectory within hull   (b) near grasping zone

(c) trajectory right, slight rotation   (d) trajectory high
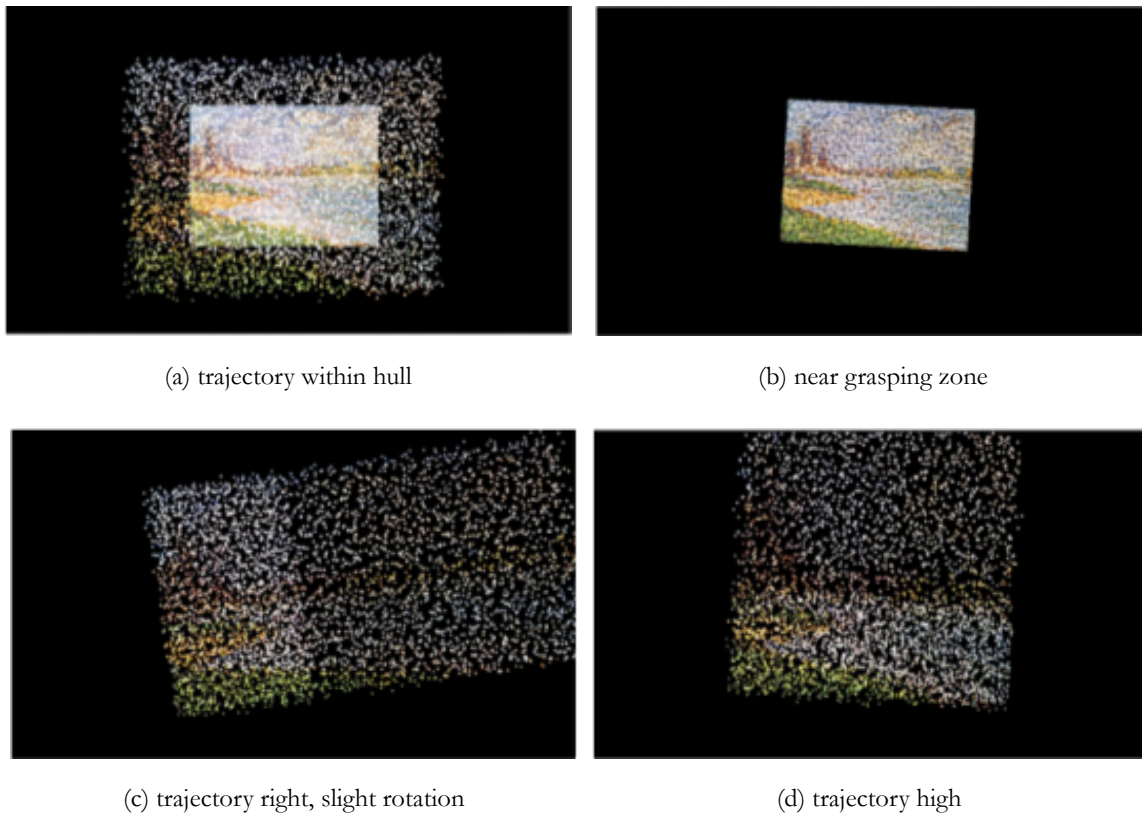
Figure 4.6: Abstract visual environment.

In contrast to the graphics engine's focus on providing spatial feedback, the music engine is focused on the temporal aspects of movement such as movement speed and smoothness. The audio feedback system is made of two programs: the first is a third-party MIDI-controlled sampler, Native Instruments Kontakt[3], which acts as our sound engine. The second is a proprietary gesture analyzer and music generator, written in Max/MSP, which processes data from the motion analysis and controls the sound engine via MIDI. The sonification engine transmits information to the user using by playing two primary orchestration elements: a solo instrument playing in the foreground, and an ensemble of instruments playing in the background. In addition, indicator sounds and audio alerts are used to minimize unhelpful compensation habits patients may have developed.

The foreground instrument may sound like a marimba, guitar, or piano. These timbres were chosen for their percussive attacks: this helps the foreground instrument be audible above the background ensemble. During each trial, the foreground plays a continuous stream of notes. To give feedback on the speed and smoothness of reaching, the rate of note production is mapped to the velocity of the user's hand. The foreground instrument plays a slower stream of equal-length notes during resting or grasping, when the hand is not moving. During reaching and returning, when the hand speed is arcing, the note-rate accelerates (see table 4.1). If the patient tends to move slowly, the acceleration will be mild or nonexistent. However, if the patient tends toward quicker movements, the acceleration can be pronounced. If patients are not moving smoothly, this mapping lets them know: tremors or jerkiness will be marked by bursts of accelerated notes for every oscillation of the movement. Ergo, jerky movement will create jerky music.

To give feedback on reaching progression, any notes that are produced are chord tones where the chord belongs to a harmonic progression mapped to depth Z of reaching and the reaching modes (see table 4.2). The chord progression is changed upon every reach for greater

---

[3]see http://www.native-instruments.com//en/products/producer/kontakt–5/

Velocity	Note Rate
0.0–0.2	♩ ♩
0.2–0.4	♩♩♩ (3)
0.4–0.6	♩♩♩♩
0.6–0.8	♩♩♩♩♩♩ (3) (3)
0.8–1.0	♩♩♩♩♩♩♩♩

Table 4.1: Mapping of reaching velocity to note-speed.

variation. During resting, the harmony remains on the first chord of this progression, and the rest of the chords are cycled through as the reaching movement is performed. In correctly-performed reaches by unimpaired users, the chords move quickly and cleanly through the progression. If users do not move with surety to the target and back, the chord progression is not smooth, meaning that the length of time spent on each chord is irregular. Or, if grasping is not achieved (i.e. the patient attempts to reach and returns, but does not reach the grasping zone) the progression is not completed. Instead, the progression reverses and cycles back to the beginning, without playing any resolution. Essentially, harmonic resolution is used as incentive for reaching the target on each attempt.

Note selection within this harmonic framework is done according to two playing styles: patterned arpeggio style or random selection. The arpeggio patterns, which are configurable by team members, specify repeated contours in the way notes are played with regard to pitch, volume, duration, and onset interval. This playing style is used when the patient is in resting mode. However, these patterns are not as musical when sped up, so during reaches, the playing style switches to random note selection, where the pitch and volume of notes is randomly selected within configurable ranges. In both playing styles, only chord tones are available for selection; this makes it impossible for wrong notes to be played.

Mode	Depth	Harmony
Resting	0.0–0.2	I
Reaching	0.2–0.5	ii
	0.5–0.8	III^7
Grasping	0.8–1.0	vi
Returning	1.0–0.3	V^7
Resting	0.3–0.0	I

Table 4.2: Mapping of reaching depth and mode to a sample harmonic progression.

The background ensemble consists of woodwind sounds (flute, clarinet, and bassoon) and string sounds (violin and cello). These sounds were chosen for their ability to blanket a wide range of pitches across the bass and treble clefs while delivering harmonic structure. To ensure the music has a unified harmonic character, the background instruments share the harmonic mappings of the foreground instrument. To ensure there is a metric framework for the music, however, the rates of background note production are not mapped to anything; they remain steady throughout the reach. Otherwise, if the note-rates of the background were mapped, the synchronized note accelerations can be perceived as interruptions to the musical continuity (this was tried in one early iteration of the system). All background instruments use random note selection throughout the reach. To give feedback on joint operation, which is an important feature to be aware of for some stroke patients, the background ensemble's volume is mapped to elbow openness. In correctly-performed reaches by unimpaired users, the background ensemble is inaudible during the resting state, but swells to full volume upon grasping, then decrescendos back to silence upon returning. Impaired users who do not fully open or close their elbow may not hear silence during resting or may not hear a loud background ensemble during grasping.

Figure 4.7 is a notated musical score from part of a good reach. The foreground instrument is marimba, and the background ensemble is enabled. This shows a chord change and a

deceleration in the marimba as the movement velocity slows, because the patient is approaching the resting zone after a reach.

Figure 4.7: Musical score generated by returning movement.

In addition to the sonification mappings that have already been discussed, the music engine also employs a success indicator, which sounds like a triangle ringing, to indicate the moment when the patient has successfully moved into grasping mode. This mapping synchronizes with: a) the chord change occurring upon grasping, and b) the coalescence of the particle field into the original image driven by the graphics engine. Also, there are three map-

pings which are designed to alert patients when they are employing unhealthy or inefficient movement strategies in the course of the reach. These three mappings are:

- Shoulder Compensation: Some patients with reduced arm strength develop a bad habit of compensating by raising the shoulder or contorting the torso. This can slow rehabilitation in the arm and lead to future health problems. If a user does this, a clattering sound fades in. The worse the compensation gets, the louder the clattering sound gets.

- Torso Compensation: Some patients with reduced arm strength develop a bad habit of leaning forward to reduce the reaching distance to objects. This can slow rehabilitation in the arm and lead to future health problems. If a user does this, a cymbal roll starts to swell. The more pronounced the leaning, the louder the cymbal roll gets.

- Trajectory Error: Supplementing the visual trajectory information (the particle field stretching) is a sonic trajectory indicator which is applied to the foreground instrument. If the patient's arm leaves the hull, the foreground instrument goes out of tune. This mapping synchronizes with the stretching of the particle field driven by the graphics engine.

The annotation, storage, and retrieval component of AMRR maintains a database, updated 100 times per second, saving the system's instantaneous state and all information transmitted between components. This includes:

- The marker location data from the motion capture system.

- All broadcasted data-frames from the motion analysis system. This is particularly useful for physical therapists, who can plot automatically-generated graphs of motion analysis data between usage sessions or between sets of reaches. The scope of these graphs can be narrowed to look at individual reaches, or widened to encompass weeks of therapy

- Any team-configurable parameter settings from the graphics, music, or motion analysis engines. These are sent to the database at the beginning of each reach.

- Notes made by team members. Team members can type observations into a web-based form in order to save them to the database. The form also contains checkboxes for frequently-made observations so that these can be entered more quickly. All annotations are visible to all team members in real-time.

- An audio/video capture system records the visuals and music generated by the system into the database. In addition, video is recorded from a camera positioned to capture the patient and video monitor (as captured in fig. 4.2). Audio from three microphones is recorded: one placed to pick up patient remarks, another placed to pick up team remarks, and a third worn by the physical therapist.

With the data saved in this database, complete usage sessions can be recalled and re-created. This component makes AMRR into a platform for scientific study on experiential media as a rehabilitative tool, and also affords longitudinal evaluations of this system's ability to rehabilitate stroke patients with right-arm impairment. These evaluations are discussed in the upcoming subsection.

4.1.2   *Evaluation*

Early in the design process for AMRR, we performed a preliminary evaluation with twelve unimpaired right-handed subjects. This preliminary test was done for proof of concept and to identify useful rehabilitation strategies. Also, we hoped to show that through careful configuration of musical parameters, we could influence the speed of user reaches. The parameters that were manipulated in this evaluation were: a) the global musical tempo affecting the speed of notes throughout the reach, b) the timbre of the foreground instrument (marimba, piano, or guitar), and c) playing style (arpeggiated verses random) during the resting state. These

parameters, and the tested configurations of these parameters, were chosen by the music experts among our team, because systematically isolating the effect of each scientific variable would not be practical in a system of this complexity.

Each test subject performed 19 or 20 sets of ten reaches, with different permutations of the three parameters on each set. We used the annotation, storage, and retrieval component to identify whether subjects hastened or slowed their movements between sets. The results of these experiments were that, although the speed of reaching invariably changed with different permutations, the direction of change was not always consistent with our hypotheses because user fatigue and impatience were confounding factors. User comments and criticisms (which were annotated in the database) commonly noted that the task, and the audio feedback, seemed tediously repetitive after a large number of reaches. This could be due in part to the unimpaired subjects' highly stylized methods of reaching. Because unimpaired subjects are capable of reaching in any manner they wish, after many reaches they may fall into a pattern of rote repetition, and this rote repetition will drive repetitive musical feedback.

Next we performed a similar preliminary evaluation with three stroke patients to see if any improvement in the features of their reaching movements could be detected after using AMRR. The stroke patients had different types and levels of impairment. For example, one had reduced arm strength and reduced ability to supinate; while another had greater strength and supination ability, but also had a deep tremor in the arm. All were cognitively sound with unimpaired language capabilities, all were right-hand dominant with right-arm impairment, and all had been healing from their last stroke for a minimum period of six months.

Figures 4.8 to 4.10 are example comparisons between an impaired subject from this evaluation and an unimpaired subject from the previous evaluation. In fig. 4.8, plotting the reaching depth over time, the stroke patient is able to reach the grasping zone, but without the accuracy and control exhibited by the unimpaired participant. The grasping movement is more tentative (i.e. the approach slows sooner and the return ramps up slower), and the

target is overshot (i.e. the plot goes higher than 1.0). In fig. 4.9, plotting reaching velocity over time, the stroke patient is not able to perform the reach as quickly as the unimpaired user (i.e. the peak of the stroke patient's graph is lower). Also, this stroke patient exhibits tremor, as indicated by the velocity oscillations. In fig. 4.10, plotting elbow openness over time, the stroke patient's elbow unfolds and folds more slowly (i.e. the plot is more rounded), and does neither to the extent that the unimpaired user's elbow does (i.e. the minimum of the graph is higher in the stroke patient, and the maximum is lower).

Each stroke patient used the session three times per week over the course of two weeks. These subjects tend to fatigue more quickly than unimpaired subjects, so usage sessions consisted of only ten sets of ten reaches. At the beginning of each usage session, system calibration was performed in such a way that these subjects could reach the grasping zone with little difficulty. Once the physical therapist judged a patient ready, the grasping zone would be reduced slightly in size and moved slightly further away. Whenever this increase in difficulty took place, another instrument was added to the background ensemble in order to reward the patient with increased musical variation.

Performing this experiment, the AMRR team began to happen upon useful rehabilitation techniques. For example, we noticed that one patient exhibited poor elbow openness during reaching. All movement was driven by the shoulder, and the elbow was not being used. We succeeded in reducing this problem through carefully manipulating the location of the grasping zone. First we moved the grasping zone a few inches to the right, so that the patient had to open the elbow in order to achieve grasping. Later, we returned the grasping zone to its original position, but by this time the patient was accustomed to using the elbow, and there was no return to the previous movement style. Patient comments proved instructive in terms of rehabilitative methodology. One patient noticed he tuned out the music because it was not in a style he enjoyed; thus he received less information from sonifications he did

(a) stroke patient   (b) normal patient
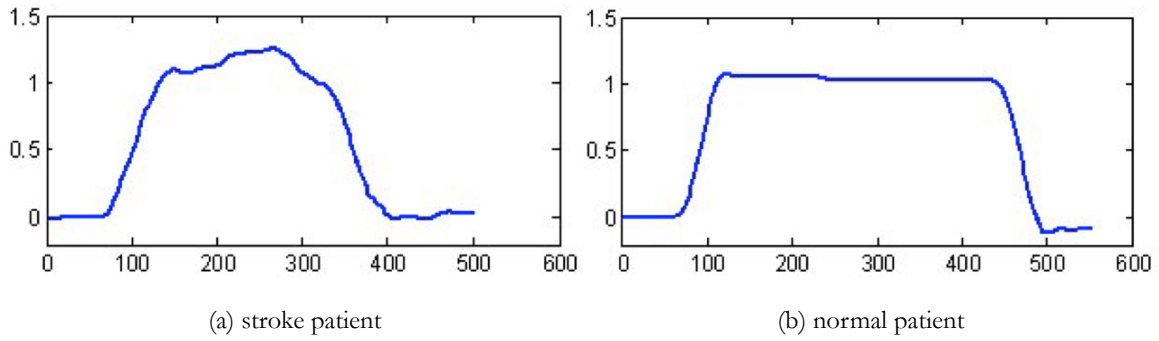
Figure 4.8: Comparison of reaching depth between an unimpaired user and an impaired user.

(a) stroke patient   (b) normal patient
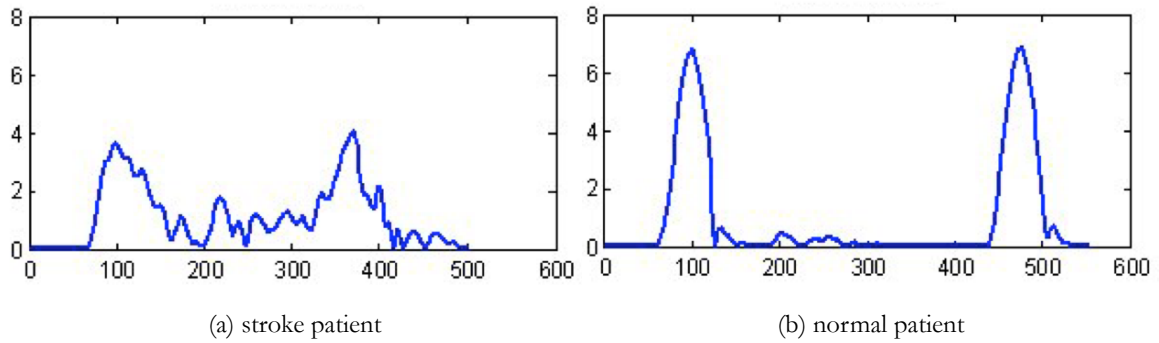
Figure 4.9: Comparison of reaching velocity between an unimpaired user and an impaired user.

(a) stroke patient   (b) normal patient
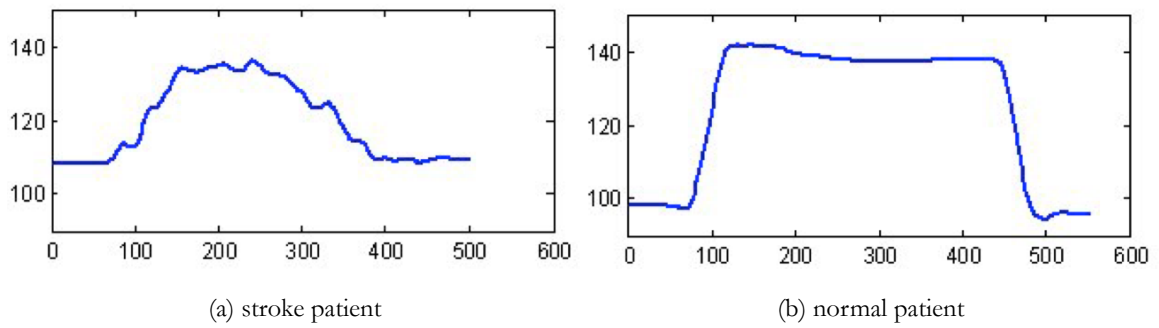
Figure 4.10: Comparison of elbow openness between an unimpaired user and an impaired user

111

not find compelling. Another patient noticed that some cartoonish imagery presented in the abstract visual environment might be seen as patronizing by stroke patients.

The results of this preliminary evaluation were encouraging. For instance, fig. 4.11 compares the last five trials of a patient's first usage session with the last five trials of that patient's last usage session. As can be seen, smoothness of reach, sureness of reach, and speed of reach completion all show improvement. Also, many observations made by the stroke patients indicated rehabilitative benefits. For example, one patient remarked, "I forgot I could still do this *[motion with this arm]*!"
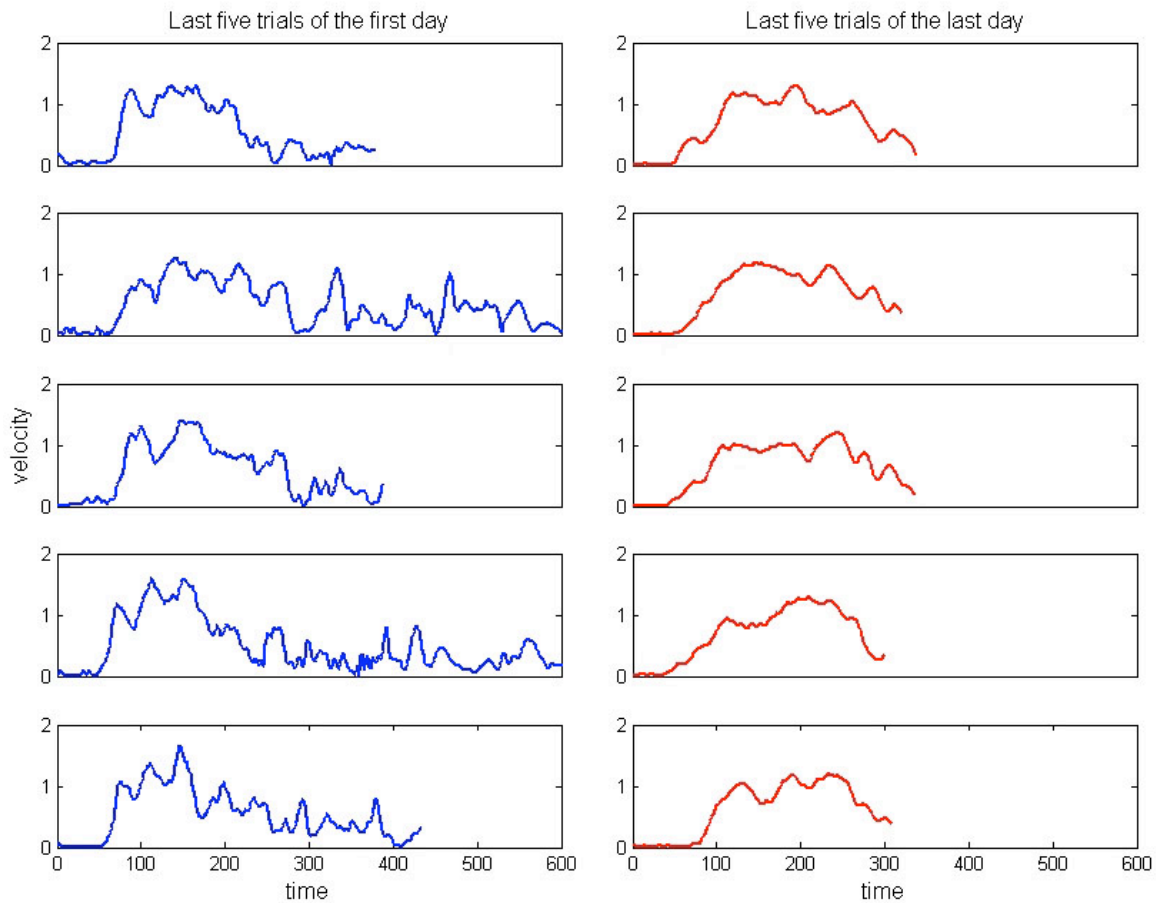
Figure 4.11: Velocity improvement of one patient: the last five trials of session 1 vs. the last five trials of session 6.

It was determined, however, that longer and more rigorous testing was required. At this point, the design of the system's music engine seemed to be in a mature state, so I left the project to pursue other research. The project continued on, and a longitudinal study was performed where a group of eleven stroke patients used AMRR three times a week for four weeks. This group's progress was compared to a control group of ten stroke patients who had an identical therapeutic schedule, but with sessions of traditional physical therapy instead of AMRR. Before and after this month of therapy, patient movement was measured using standard measurements of impairment such as:

- The Wolf Motor Function test: where a therapist rates, on a 0 to 5 scale, a patient's ability to perform a reaching task within two minutes. (Wolf et al., 2001)

- Upper Extremity Fugl-Meyer Assessment: where a therapist rates the patient's affected arm joints with regard to range of motion, pain, sensation, and proprioception; and also rates the patients motor function in the arm. All ratings are combined to result in a general impairment rating on a 0 to 126 scale. (Duncan et al., 1983; Fugl-Meyer et al., 1975)

- Motor Activity Log: where patients self-report, on a 0 to 5 scale, how often they use the affected arm in their day-to-day life. (Uswatte et al., 2005)

- Stroke Impact Scale : where patients self-rate the impact the stroke has had on areas of their lives such as social interaction, emotion, cognition, and motor function. Ratings are combined to result in a normalized scale from 0 (no recovery) to 100 (full recovery). (Duncan et al., 1999)

In addition to these standard tests, a set of measurements called Kinetic Impairment Measures (KIMs) were designed to use motion capture to rate certain specific qualities of patient movement. KIMs depend on a corpus of reaching data created by motion-capturing the

reaches of people at varying levels of impairment due to stroke (including unimpaired persons who have not had a stroke), where the people are asked to perform two specific reaching tasks. The first task is reaching to grasp a cone, as all AMRR users do in the course of calibration. The second task is reaching to press a button, which is not part of AMRR therapy. KIMs are calculated by asking stroke patients to perform these same tasks under similar conditions, then comparing the reaches with the corpus to obtain a zero-to-one value where one means the patient's reach is similar to the reaches of impaired users, and zero means the patient's reach is similar to unimpaired users. KIMs measure a reaching task that AMRR users do not train for—reaching to press a button—and this helps ascertain whether AMRR rehabilitation generalizes into forms of patient movement outside the specific reaching movements that AMRR users practice. KIMs can be calculated with regard to any movement feature. As listed in Duff (2012), the KIMs used for this evaluation were:

- Trajectory: horizontal and vertical efficiency and consistency of reach;
- Targeting: error and consistency of hand placement during the manipulation phase of the reach;
- Temporal: peak reach speed magnitude, consistency, and time to target;
- Velocity: smoothness of reach and how well the velocity curve adheres to an ideal bell shape;
- Compensation: excessive shoulder or torso movements during the reach;
- Joint Function: shoulder, elbow and wrist range of motion and consistency;
- Joint Correlation: synergy between key joint pairings of the arm such as shoulder and elbow.

The results of this evaluation show that AMRR is effective at improving some aspects of the reaching movement of stroke patients. With regard to the KIMs, patients who used AMRR showed significantly greater improvement than patients from the control group. AMRR users showed significant improvement in all of the movement qualities measured by KIMS except targeting, while the control group did little to improve ratings in any of the KIMs except trajectory. A composite KIM score consisting of a weighted combination of all other KIMs showed

significant improvement only in AMRR users. With regard to the more standard assessment measures, Duff (2012, pg. 97) claims:

> Both the AMRR group and Control group significantly improved their WMFT FAS score. Neither group had a significant change in their median time to completion of the WMFT tasks. Both groups increased their FMA Motor Function scores significantly, but only the Control group significantly increased the scores in the Range of Motion/ Pain/Sensation sections. Neither group had a significant increase in the MAL for either the Amount of Use or Quality of Movement sections. The Control group significantly improved their Stroke Impact Scale score from a reported 64.41% recovery to a 76.48% recovery. Between group analysis of the FMA scores shows that the Control group's increase was significantly greater than that of the AMRR group in both sections.

One previously-overlooked benefit of this system emerged from this evaluation, namely, how this system functions as a tool for training physical therapists. The system was designed to be highly adaptable in order to afford quick changes to the therapeutical protocol at the will of the attendant physical therapist. This allows therapists to respond to day-to-day changes in patient movement and abilities and deploy new therapeutic strategies between therapy sessions. The composite KIMs tell an interesting story: only two AMRR users had composite KIMs that went down over the course of the four weeks of therapy, and these were the first two in the study. Similarly, the two patients who had the best overall improvement in composite KIMs were near the end of the study. This suggests that our physical therapist may have changed approaches as the study progressed in order to better leverage the AMRR system to address the dynamics of the reaching movement in impaired patients. This in turn suggests AMRR may be useful as a tool for training clinical physical therapists how to address these dynamics of reaching.

4.1.3   Analysis

According to the framework, AMRR should draw volitional usage in the shorter-term, and my recollections of user reactions support this conclusion. Promoting the system to the target user base was accomplished through a particularly powerful form of word-of-mouth:

we asked local doctors to spread the word and help us recruit stroke patients fitting evaluation criteria. Many stroke patients are intensely interested in recovering their abilities, so AMRR's stated application could not be better targeted to the user base. The system location is inconvenient, but not more so than the regular physical therapy sessions these patients are already used to attending in the clinic several times per week. Motion-capture based rehabilitation systems are not in common usage yet, so AMRR would seem novel to the user base at first, and since stroke patients tend to be older, many would be unfamiliar with multimodal interactive systems in general. There are many elements of AMRR detracting from usability, but team members were present for all usage sessions and could instruct users through points of confusion. AMRR usage sessions are designed, from the calibration step forward, to ensure that the reaching task stretches the capabilities of impaired users, so this should help ensure the system remains cognitively rich for stroke patients.

On the other hand, according to the framework, AMRR is not as likely to draw volitional usage in the longer term, and my recollections also support this conclusion. Although incrementality is well-designed in AMRR, no other heuristic for longer-term volitional usage is. For one example, immediacy is poor: users can not run the system by themselves, a team of experts is needed; plus 15–20 minutes of prep time is required (placing markers, calibrating motion capture, and so forth) before a usage session can commence. For another, demonstrability is poor: users are asked to perform in front of an audience of unimpaired strangers (consisting of the AMRR team), knowing the task they are struggling with would seem trivial to these observers. By far, however, the most problematic aspect of AMRR is its explorability.

Movement rehabilitation systems such as AMRR pose particular design challenges with regard to explorability. AMRR, like most interactive movement rehabilitation systems, belongs to a category called 'task-based systems,' modeled after traditional physical therapy approaches, where patients repeat simple tasks that are frequently needed in day-to-day life. The hope is that patients will improve with regard to these tasks, and the improvements will generalize into

other areas of patient movement. Unfortunately, this approach borrows the quality of rote repetition that is present in traditional physical therapy exercises, and this rote repetition is part of the reason why patients do not always comply with physical therapist directives regarding exercise at home. That is, many patients want to recover, but cannot bring themselves to do the exercises unless a physical therapist is supervising, because the exercises are tedious activities.

Task-based rehabilitation depends on the idea that there is a 'normal' way to move and an 'impaired' way to move. There is a solid medical basis for this dichotomy. Impaired patients frequently develop compensation strategies: styles of movement working around the impairments. From the standpoint of rehabilitation, these compensation strategies are usually detrimental. At the least, they result in less practicing of movements that should be practiced. Consider patients with reduced arm strength: if they develop compensation strategies that allow going through life without using the weakened arm, then the muscles will never be exercised back to full strength. At worst, these compensation strategies can result in future health problems. Consider patients with one sore knee: compensation strategies developed to ease the pain on the sore knee may put strain on, and eventually lead to problems in, the other leg. Therefore, rehabilitative systems should discourage these compensation strategies whenever possible.

Because of the rote repetition (which is antithetical to complexity) and because of the need to discourage compensation strategies (which is antithetical to operational freedom), task-based systems are not conducive to explorability. In the case of AMRR, the system's task is reaching forward for the purpose of grasping. Users are asked to practice this motion dozens of times as correctly and consistently as possible. If they do this, the feedback will be similar each time. Even though the feedback is designed to be pleasant on correctly-performed reaches, this self-similar feedback becomes tedious over many repetitions. The paradox is, if users seek to alleviate the tedium by exercising more operational freedom in their reaching movements,

117

the system is designed to correct them (i.e. alert them to their behavior) by playing less pleasant feedback.

As it happens, the designers of AMRR hoped it would be used over longer-duration time frames, and some of the evaluations in section 4.1.2 entailed usage careers that were weeks long. If AMRR is poorly designed for longer-term usage, how were these evaluations accomplished? The only way they could have been—through the use of extrinsic incentives. First, all evaluation subjects were paid. Second, in light of the fact that many insurance companies will cover only six months of physical therapy, all participating stroke patients received free physical therapy. Third, part of the the attending physical therapist's job was to help patients stay on task—using direct supervision, encouraging speech, badgering, and if necessary, physically manipulating the patient's arm—whether they felt engaged or not. Lastly, circumstances themselves may provide a feeling of extrinsic necessity, having many similarities with an incentive structure, to patients. Rehabilitation is a serious matter and patients know that recovery is important to future well-being. If there had been no stroke, many of these patients could not have been induced to participate in these evaluations, but the prospect of fuller recovery may have made them feel compelled. Although these extrinsic motives drive users to participate in many usage sessions, these are not *volitional* usage sessions. It is unlikely that many users would continue to use this system of their own volition without outside incentives.

If I were re-designing the stroke project from scratch, I would abandon task-based rehabilitation in favor of making a practice-oriented system. I would attempt to design a system that is a hobby in itself, one that is easier and more fun to do if done with good movement form. There are precedents for this hobby-centric type of rehabilitation. For example, learning a new musical instrument is known to alleviate depression for many people, and seeing that instrument practice demands movement, instrument learning has been used to help sufferers of Parkinson's Disease improve their symptoms of bradykinesia (movement slowness) and depression at the same time (Pacchetti et al., 2000). If patients take up hobbies that are

rehabilitative, then rehabilitation becomes fun, and depression and apathy become less likely. Thus, the idea genesis for the system discussed in chapter 5 can be traced to AMRR; it was originally designed to be one of these practice-oriented movement rehabilitation systems.

In my re-designed stroke system, I would eschew extrinsic motives of all kinds. Because of the tendency for rehabilitation to feel like a necessity, I would favor an aesthetic based on creativity, light-heartedness, or playfulness; because, paradoxically, I would favor making system usage seem less important. This would, in turn, allow patients to feel that their usage is voluntary rather than mandatory. The version of AMRR I worked on is not creative, playful, or light-hearted. Using it looks and feels like being integrated into a futuristic machine:

- The experience begins with people affixing optical markers all over the patient's right side.

- Then, patients are placed in a sliding table-and-chair contraption that is difficult to get out of.

- In order to calibrate AMRR, patients are asked to manipulate objects in specific ways.

- Once all these preporatory steps have been taken, patients then perform calculated series of reaches as consistently and regularly as possible so that AMRR will provide them with consistent and repetitive feedback.

- If patients do not or can not perform reaches in the optimal way, AMRR will correct them.

- All this is done under the watchful eye of a team of several researchers and a physical therapist.

This activity could not be designed to be more intimidating, uncomfortable, clinical, or detached. In order to address this problem, the AMRR team integrated arts into the feedback. It was hoped that using orchestral music in the sonifications and visual images in the graphics would ameliorate the machine-like quality of the system. These artistic elements are the only parts of AMRR designed to impart variation rather than rote repetition. They also afford certain emotional induction strategies. For example, resolution (as discussed in section 2.3) is used to provide a reward for successful reaching. The particle field is designed to create an unstable state; re-forming that particle field into the original image should return the scene to

119

stability, thus bringing a sense of relief. Similarly, the harmonic progression provides resolution (by moving in expected ways from unstable chords to stable chords) if successful grasping is accomplished.

In my opinion, the arts integration that was implemented was not sufficient to overcome the aesthetic challenges posed by the system's form. The system remains unexplorable, and continues to create a perceptibly daunting emotional atmosphere. Users of the system often seemed either bored or frustrated, and sometimes poor movement performance (with regard to tremor, for example) was attributed to nervousness. In many ways, the inspiration of my dissertation topic can be traced to the dilemma discussed in this subsection: that AMRR is designed for and requires long-term usage for success, but is not engaging past the short term. Many of the other systems discussed in this chapter were designed to test ways of addressing issues encountered while working on AMRR. For example, the system described in section 4.2, which was designed to explore ways of handling mood in interactive media, was a direct response to the tension between two facts about AMRR: a) AMRR patients seemed to exhibit less tremor and more movement control when calm, and b) AMRR itself sometimes seems to induce a feeling of nervousness.

Many of the approaches discussed in chapter 2 are supported by experience I attained by working in AMRR. For example, designing musical sonifications for AMRR led me to embrace the channel model of modalities discussed in section 2.3. As discussed there, music is a modality that is capable of transmitting information, and designing interactive music to transmit information is the purpose of musical sonification. AMRR users, however, received much less information from the music than was encoded in it, not because they could not hear or understand said music, but because they simply tuned it out. Their cognitive load was such that it was simply not possible for them to respond to any musical features beyond one or two of the most obvious mappings. I soon realized that it is difficult for people—even for music specialists like myself—to actively listen to more than one musical feature at a time.

The channel model of modalities, since it implies a rule of cohesion, explains why: more than two simultaneous streams of information within a modality will surely overwhelm the available bandwidth. This makes it a highly challenging task to sonify the complexities of human movement. For example, in AMRR, movement was detected in three dimensions, so describing that movement requires a sonification mapping for movement in X, Y, and Z, and this violates the rule of cohesion.

From the rule of cohesion emerges an omnipresent tradeoff between a sonification's understandability versus its informativeness. If a sonification must be understandable, then the number of mappings must be reduced to one or two. If a sonification can be confusing at first, then more mappings are allowable; over time, listeners may acquire skill at parsing the music, and then their confusion will be reduced. If sonification mappings are intuitive, meaning they leverage commonly held mental-metaphorical relationships (e.g. movement speed is mapped to musical tempo), this may hasten the acquisition of this parsing skill. Even with skilled listeners, however, the sonification mappings will need to be sonically distinct and cleverly-designed. Here I give a set of descriptors for sonification mappings. These descriptors may help sonification designers select appropriate mapping styles for the types of information-streams being transmitted:

- **Consistent and Optimized Mappings:** Mappings which are consistent, meaning they always respond in exactly the same way to the same user movements, are more learnable. Mappings which are optimized, meaning that a audible result is produced for the type of movement the system is exploring, are more understandable. In most systems, mappings should be both consistent and optimized at the same time. The exception is this: it is possible to design task-based systems where users must switch between multiple tasks. In this case, it is no longer possible to optimize the mappings for each task while maintaining mapping consistency; therefore a choice must be made between learnability and understandability.

121

- **Descriptive verses Comparative Mappings:** Descriptive mappings exist merely to make the listener aware of structural features: in these mappings the musical element will track the contours of the underlying data. Comparative mappings exist to make the listener aware of a deviance—often a deviance from some ideal—thereby alerting the listener that adjustments need to be made. Comparative feedback is not to be confused with negative reinforcement, as it is possible to make comparative feedback a pleasant form of information transfer.

- **Continuous verses Discrete Mappings:** Continuous mappings are pervasive streams of sonic information which are constantly transformed by sonification mappings. Discrete mappings only appear when needed. Discrete indicators may be one-off sounds which are triggered by certain events (i.e. alerts), or they may be partially continuous, fading in only when needed and fading back out later on.

The channel model implies that the rule of cohesion can be circumnavigated by dividing the information-streams among different modalities. For example, if some streams of information can be transmitted via text, speech, or graphics instead of music, all of the feedback may become more understandable. The mapping styles listed above may help classify the information-streams in the minds of listeners, so listeners can treat the different mapping styles as if they were different modalities. If so, one strategy for creating understandable complex sonifications would be to employ a diversity of these mapping styles. This was the approach taken in AMRR.

The channel model explains, in part, why the repetitiveness of the reaching motion is so detrimental to engagement: it violates the channel model's rule against redundance. Let us perform a thought experiment: let us imagine that AMRR were redesigned so that, although the reaching task were left the same, the feedback given was suddenly more dynamic and interesting. For example, let us imagine that the visual and auditory feedback were redesigned to be as

engaging as that seen in video games, but still controlled by performing the reaching movement as consistently and repetitively as possible. Would this make AMRR engaging over the long term? In my opinion, it would not. The reaching movement, according to the channel model of modalities discussed in section 2.3, is redundant. Performing the same movement over and over is annoying because it asks people to expend the cognitive effort of using the kinetic and proprioceptive modalities, without using much bandwidth from either. This is wasteful effort, frequently resulting in a feeling of annoyance.

In this subsection, I have been critical of AMRR. It is important to note, however, that our evaluations suggest that AMRR patients receive real rehabilitative benefits from usage. Part of the reason for this is that AMRR is well-designed with regard to the heuristic of incrementality; every increase in user skill is quickly accounted for so that a slightly increased challenge level can be delivered. During the testing with unimpaired subjects, we noticed that AMRR had an interesting learning curve because it requires skills that are unneeded by unimpaired people in day-to-day reaching. As users are physically reaching for targets that exist only in virtual space, they must either: a) use the feedback to narrow down the location of the grasping zone, or b) remember, proprioceptively, how it felt to successfully reach the grasping zone on previous reaches. In day-to-day reaching, it seems that most people know how reaches will be done before-hand, and little course correction is needed. It takes practice, therefore, to learn to adjust mid-reach according to real-time immediate feedback. Beginning users must painstakingly search the space with their arms to find the grasping zone, but experienced users, having mastered the skills of adjusting mid-reach and consciously using proprioception, can find the grasping zone quickly even if the AMRR team moves it without notice. Although these skills are not frequently needed by unimpaired people in day-to-day reaches, it is possible that they are key for many stroke patients in speeding the recovery of movement faculties.

After I left the project, the AMRR team continued by designing and implementing a home-based version of the system. The team as a whole shared many of my concerns regard-

ing long-term usage in AMRR, and their work on the home system attempts to address many of these issues. By locating the system in the home, they improve its convenience and they greatly reduce some of the previously-discussed problems with regard to system demonstrability. In order to achieve a home-based installation, they also improved system cost and ease of both setup and usage. These enhancements, along with many other changes that were made, should increase the longer-term volitional usage in this system. Lehrer et al. (2011a,b) provide discussions of this home-based system and the reasoning behind many of its design decisions.

4.2   Emotional Slideshow Generator

As discussed in section 4.1.3, AMRR carried some unintended emotional connotations. This emotional slideshow generator (ESG) was a response to this problem; it is designed as a case study on how to handle mood in multimodal systems. In day-to-day life, we see that mood is transmitted via multiple modalities in order to induce immersion and cause viewers to become emotionally invested in story-lines. For example, emotional music seems to be an important element in storytelling media such as television or film, and is often used to support emotional imagery, dialogue, and narrative. The same mood is being presented in multiple modalities; according to the the channel model rule of synchronization (as discussed in section 2.3), this drives home the mood more emphatically and more clearly. ESG was designed to explore and study this type of synchronization, while removing the pre-composed narratives, in order to identify design elements needed for immersion and for transmission of mood.

ESG automatically compiles and flips through a slideshow, using images found on Flickr and downloaded in real-time, with a user-selected mood. While this goes on, background music is played that delivers the same user-selected mood. This system was designed and implemented by myself, with guidance from Prof. Todd Ingalls and Dr. Ellen Campana. By far, the musical component of this system took the most research and development time. Unlike the slideshow component which compiles existing content, the musical component generates emotional musical compositions from scratch. Wallis et al. (2008) describes the design of the musical component of the system, and an evaluation of that emotional music synthesis, assisted by Janel Goodman, is given in Wallis et al. (2011a). Some language from these publications is incorporated here. ESG can be considered an interactive artwork that is designed to be exhibited during talks, or in classroom settings.

4.2.1   Description

This system is software-based, requiring a computer attached to a visual display and sound system. A fast internet connection is required because ESG downloads images from Flickr, a web-site designed for social image sharing—at the time of ESG's creation, Flickr's database of images was larger than any of its competitors[4]. Flickr provides an API affording automatic image searching (via tagged metadata) and downloading. Users select a mood from a drop-down menu (see the interface in fig. 4.12), whereupon the slideshow generator searches Flickr for images that have been tagged with words related to the selected mood. For example, if the user selects 'angry,' Flickr will be searched for images tagged with words containing the word 'angry' along with related words such as 'anger,' 'rage,' 'furious,' and so forth. Once images are found, they are automatically downloaded and displayed as slides.

Figure 4.12: Emotional slideshow generator interface.

---

[4]see http://www.flickr.com/

Many images are invariably found, creating a backlog, and this backlog of images allows ESG to modulate the speed of slide-switching according to the arousal of the desired mood. The arousal values used are given in table 4.3—these values were calculated using data from Russell (1980), where subjects classified 28 emotion words according to the approximate degree heading on a circumplex model of emotion. Table 4.3 was calculated by taking Russell's list of ratings, finding the mean rated heading for each word, then finding the cartesian coordinate of that heading on the unit circle. Since, as discussed in section 2.3, speed is a universal emotional feature mapped to emotional arousal, quicker slide-switching is used for moods with high arousal (i.e. anger, joy), while slower slide-switching is used for moods with lower arousal moods with low arousal (i.e. serenity, sadness). The transition speed and effect is also mapped to the arousal of the selected mood. In low-arousal moods, slides transition through slow fading. In high-arousal moods, the slides transition through quick cutting or through transition effects such as page-turning.

The valence and arousal values of the selected mood are also used to control numerous features of the music generation algorithm, which algorithmically generates piano music to fit the desired mood. As seen in fig. 4.13, the algorithm contains three structural modules implementing ten musical features, where those features are mapped to valence and arousal according to table 4.4. Each parameter is linearly mapped to valence or arousal in the most continuous possible way; for example, harmonic mode thresholds the valence selection space into six equal regions, while loudness increments between low and high arousal settings. The algorithm is designed so that musical parameters can be easily re-scaled or re-mapped, so the values in table 4.4 represent the algorithm in only one configuration. In practice, often a scaled-down subset of parameters was used, in which the voice spacing and voice leading parameters were fixed at a central value across all valence settings. In preliminary study, these simplifications seemed to have little effect on the emotional connotations of the algorithm. For example, tempo was often fixed at 80 bpm across all arousal settings; the rhythmic roughness parame-

Emotion	Valence	Arousal		Emotion	Valence	Arousal
Afraid	-0.41	0.76		Excited	0.63	0.71
Alarmed	-0.1	0.86		Frustrated	-0.56	0.43
Angry	-0.12	0.74		Glad	0.87	-0.16
Annoyed	-0.43	0.53		Gloomy	-0.78	-0.42
Aroused	0.26	0.88		Happy	0.84	0.11
Astonish	0.31	0.85		Miserable	-0.89	-0.14
Bored	-0.39	-0.69		Pleased	0.85	-0.1
Calm	0.68	-0.65		Relaxed	0.68	-0.62
Content	0.76	-0.55		Sad	-0.72	-0.2
Delight	0.77	0.36		Satisfied	0.73	-0.61
Depressed	-0.72	-0.41		Serene	0.77	-0.47
Distressed	-0.67	0.56		Sleepy	0.03	-0.95
Droopy	-0.2	-0.86		Tense	-0.04	0.81
Ease	0.71	-0.59		Tired	-0.05	-0.96

Table 4.3: Valence-Arousal values for emotion words, calculated from circumplex model data given in Russell (1980).

ter alone was sufficient to affect the perceived rapidity and arousal of the music. In the rest of this subsection, I describe the meaning of the musical parameters referenced in fig. 4.13 and table 4.4.

Rhythm & Timing   As seen in fig. 4.13, the rhythm and timing module implements three parameters that are each mapped to arousal: tempo, rhythmic roughness, and articulation. The music generator creates music with a $\frac{4}{4}$ time signature, signifying that the quarter-note is the basic unit of musical time, and measures are four beats long. All note-lengths are defined in relation to the quarter-note, so scaling the duration of quarter-notes speeds up or slows down
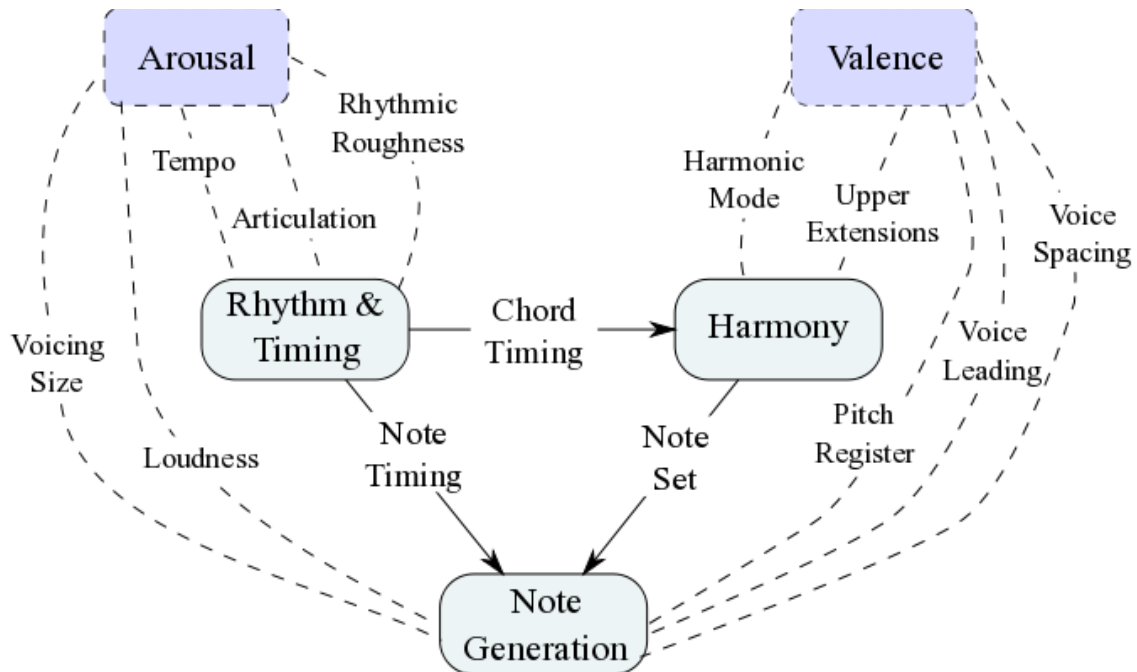
Figure 4.13: ESG musical synthesis structural modules and composition (shown by solid edges), the musical parameters they implement (shown by dashed edges), and the mood descriptors (valence or arousal) those parameters are mapped to.

the speed of all generated music. Thus the standard usage of musical tempo is a global speed value defined by assigning a duration value (in beats-per-minute) to the quarter-note. When tempo is used in ESG, it is mapped in an exponential way so that the arousal midpoint produces a tempo of 85 bpm, the arousal minimum produces a tempo of 60 bpm, and the arousal maximum produces a value of 120 bpm. Due to the naturally logarithmic perception of tempo by humans, this exponential mapping should produce a more even response to interaction; that is, equally-spaced values above or below the arousal midpoint should produce equal-sounding tempo shifts above or below 85.

Rhythmic roughness determines the internal rhythm of generated measures. The standard usage of rhythmic roughness, as determined by music psychologists such as Gundlach (1935), denotes the extent to which note-lengths are similar within a musical phrase. If all notes are of identical length, roughness is low, but if note-lengths vary widely, roughness is high. Therefore, in ESG, measures generated at minimum roughness contain nothing but six-

	Valence Min	Valence Max
Mode	Phrygian	Lydian
Extensions	2–4 added pitches	0 added pitches
Pitch Register	Centered on C4	Centered on C6
Voice Spacing	Avoid 2nd interval	Avoid 6th interval
Voice Leading	Within 3rd interval	Within 5th interval

	Arousal Min	Arousal Max
Roughness	9 rhythmic units joined	0 rhythmic units joined
Tempo	♩ = 60 bpm	♩ = 120 bpm
Articulations	2x note-length	0.5x note-length
Loudness	Velocity 50	Velocity 70
Voicing Size	2-note chords	8-note chords

Table 4.4: Musical feature mappings to valence and arousal in ESG.

teenth notes. As roughness increases, pairs of adjacent notes are selected at random and joined into longer notes. Notes that have previously been joined may be selected to be joined again, thus allowing notes of many different lengths. The following example rhythms illustrate this process of moving from smooth to rough:

Smoothest (16th notes only):

Rougher (4 notes joined to adjacent notes at random):

Roughest (5 more notes joined, 9 total):

This implementation of rhythmic roughness, therefore, also controls note density, because smoother rhythms have more notes per measure than rougher rhythms. Higher note densities are perceived as faster, so roughness is mapped to arousal so that higher arousal val-

ues produce smoother rhythms and vice versa. In ESG, each generated measure takes the rhythm of the preceding measure unless the rhythmic roughness parameter is changed.

The note-lengths given by rhythmic roughness determine the onset of notes, and they allot time for those notes, but the articulation parameter determines how much of the allotted time is actually used. When arousal is high, notes use only half of the allotted time. In music terminology, this is staccatto articulation; notes sound short and sharp. At the arousal mid-point, notes use all of the allotted time. In music terminology this is legato articulation; notes smoothly run together. At the arousal minimum, notes use double the allotted time, meaning they overlap. In music terminology, this is sostenuto articulation, which can only be achieved in certain instruments. For example, on pianos, sostenuto articulation is achieved by pressing the sustain pedal.

Harmony   As seen in fig. 4.13, the harmonic module implements two parameters that are each mapped to valence: upper extensions and harmonic mode. Discussion of the ESG's harmonic module requires specialist language from the field of music theory. Refer to Persichetti (1961) for definitions of terminology or for more in-depth explanation of music theory concepts.

The harmonic module receives a signal from the rhythm & timing module at the beginning of each measure. This signal triggers chord changes. ESG takes, as a configuration, a pre-defined chord progression written to sound good in the ionian mode (also known as the major mode or major key). One frequently-used progression is ‖: I | vi | IV | V :‖. These roman numerals denote the root notes of chords, and this sequence of chords is played cyclically and repeatedly at a rate of one chord per measure. ESG uses tertiary harmony, meaning its chords are constructed of root notes plus notes incrementing upward by intervals of a third. The simplest chords, called triads, consist of three notes: the root, the 3rd above the root, and the third scale degree above that (which is a 5th above the root). Chords can be inverted, mean-

ing that notes can be transposed upward or downward in increments of an octave (ESG uses seven-note diatonic scales, meaning that octaves are equivalent to 8th intervals) without changing the fundamental characteristics of the chord. This means that the roots of chords are not necessarily the lowest-pitched notes produced by ESG.

The harmonic mode parameter switches the underlying mode that the chord progression is based on. Through a process of rotating the ionian mode—i.e. transposing the root note of the scale up an octave and treating the second note as the root of a new mode—seven natural diatonic modes can be created. In order, these modes are: ionian, dorian, phrygian, lydian, mixolydian, aeolian, and locrian. Each of these modes has a distinct and recognizeable musical quality. According to Persichetti (1961), these modes can be sorted according to their musical brightness, and doing so results in the following ordering, from brightest to darkest: lydian, ionian, mixolydian, dorian, aeolian, phrygian, and locrian. ESG uses six of these seven modes, and maps them to valence in order of brightness so that the highest valence produces music in lydian and the lowest valence produces music in phrygian. Locrian is excluded because it is rarely used in traditional classical music or modal jazz. In all modes, the first chord serves a tonic function, and music is perceived by listeners as belonging to a certain key and mode when harmonic resolutions end in the tonic. Since harmonic resolutions depend on chords of relative instability moving to chords of relative stability, locrian is an intractable mode; its tonic chord is the most unstable triad available, so no other chord from the mode can precede it to form a harmonic resolution.

In diatonic harmony, the triads built upon different scale degrees serve different functions, but these can be grouped into three functional categories: tonic (represented in the ionian mode by the I, iii, and vi chords), subdominant (represented in ionian by the ii and IV chords), and dominant functions (represented by the V chord or, occasionally, the vii°). These three functions, due to their interrelationships, carry specific implications with regard to the structure of chord progressions. For example, dominant-function chords, as chords of high instability,

132

tend to resolve to tonic-function chords. Different modes contain different configurations of functions; and this complicates the handling of chord progressions in ESG. One cannot change the mode of a chord progression merely by using the new mode to build triads on the same roots. The musicality would suffer, there would be little perceptible relationship between the two chord progressions, and in many cases, the resulting progression would be perceived by listeners as belonging to some other mode than the intended. Instead, ESG substitutes chords from the original progression with chords from the new mode serving the same function. This results in chord progressions that are perceptibly similar to the original pre-configured chord progression, but that also sound like they belong to the new mode.

On any triad within a mode, there are somewhere between two and four notes—forming 7th, 9th, 11th or 13th intervals with the root—which can be played without forming a minor 2nd interval, or any combination of octave transposition and a minor 2nd interval, with any of the three triadic notes (the minor 2nd is highly discordant). These extra notes are called upper extensions, and when the upper extensions parameter increases, more of these notes are allowed into the chords played by the EMS system. Allowing these upper extensions into the harmony has the effect of increasing harmonic complexity and ambiguity. Seventh chords and higher can be interpreted in more than one way. For example, adding the 7th means a chord can be interpreted by listeners in two ways: as if it were based on the original root triad, or as if it were based on the second note of that triad, the 3rd scale degree. If upper extensions beyond the 7th are added, even more ambiguity due to potential harmonic interpretations will exist.

Note Selection   ESG depends on the standard MIDI synthesizer that is built into Apple computers. This synthesizer, in its default configuration, generates piano sounds upon receiving MIDI note-on messages. The note selection module interfaces with this synthesizer by sending MIDI note-on and note-off messages to it. As seen in fig. 4.13, this module implements five parameters, two of which are mapped to arousal (voicing size and loudness) and

133

three of which are mapped to valence (pitch register, voice leading, and voice spacing). The note selection module receives a signal from the rhythm and timing module for each rhythmic note, and this triggers a specific number of MIDI note-ons, determined by the voicing size parameter. At minimum arousal, two notes onset per rhythmic event. At maximum arousal, eight notes are generated. The MIDI velocity of each transmitted note-on message is stochastically determined by the loudness parameter. The loudness parameter defines the center of a random number generator that determines the MIDI note velocity of each generated note. The numbers produced by this random number generator fall into a gaussian distribution with a standard deviation of twelve. Therefore, at minimum arousal this generator produces note velocities that vary, but which center around a velocity of 50. At maximum arousal, the note velocities center around a velocity of 70. On a perceptual level, these parameters interact, because playing more simultaneous notes at the same velocity results in a louder chord.

The pitch of generated notes is determined by a rule-based interaction between the pitch register, voice leading, and voice spacing parameters. The rules can be described as follows:

- Rule 1—Pitch Register: Notes are selected from a three-octave range of chord tones, where the center note of the range is determined by the pitch register parameter. Notes tend to be selected from the center of this range unless there are conflicts with other note-selection rules precluding this. This rule keeps the music from going too high or too low.

- Rule 2—Voice Spacing: Piano players frequently modulate the overall timbral and harmonic quality of the music by modulating the openness or closedness of their chord voicings. In closed voicings, chords are played with little intervallic distance between notes, and in open voicings chords are played with greater intervallic distance between notes (voicing openness is, of course, physically limited by the anatomy of the pianist's

134

hands). To emulate this in ESG, new notes tend *not* to play within a specific interval, determined by the voice spacing parameter, of notes that are currently on. The relative likelihood of a new note being played is reduced further and further as its potential interval with playing notes becomes smaller. On a piano, one cannot play two simultaneous instances of any note, because only notes that are in the up position can be pressed down. This rule, therefore, has the added effect of disallowing that.

- Rule 3—Voice Leading: Pianists often follow a principle called voice-leading where new chords are voiced to be as similar as possible, in terms of the intervals and placement on the keyboard, as previous chords; this minimizes the required arm and finger movement. Switching from one chord to another requires finding an inversion of the second chord with notes that are proximally-located to the notes of the first chord. To emulate this in ESG, new notes tend to play within a specific interval, determined by the voice leading parameter, of notes that were recently released.

Tensions between these rules can lead to complex but human-sounding piano emulations. For example, if the voice leading interval is set narrow and the voice spacing interval is set wide, the result is directional music gestures: new notes will tend to play increasingly high or increasingly low until a conflict with rule 1 demands that the musical gesture is re-centered.

Evaluation of Music System   To evaluate the emotional music synthesis algorithm, we recruited eleven participants—five female, six male—from a student pool at ASU consisting of psychology students, whose ages ranged from 18 to 21 years of age. As per the requirements of recruiting from this student pool, each participant was given class credit in return for participation. Each participant was stationed at a computer terminal consisting of monitor, keyboard and mouse, and speakers. Several practice trials were completed, and when the participant felt ready the experiment began. Each trial proceeded as follows: First, the EMS algorithm generated music at a specific valence and arousal setting—randomly selected from 36 valence/

135

arousal configurations evenly dividing the parametric space into a 6x6 grid—for 15 seconds. Once the music stopped, a clickable grid appeared on the graphical user interface (see fig. 4.14) and the participant clicked on the valence-arousal location most closely matching, in his or her opinion, the generated music. After the click, the grid disappeared and another trial began. Each participant completed four blocks of 36 trials.
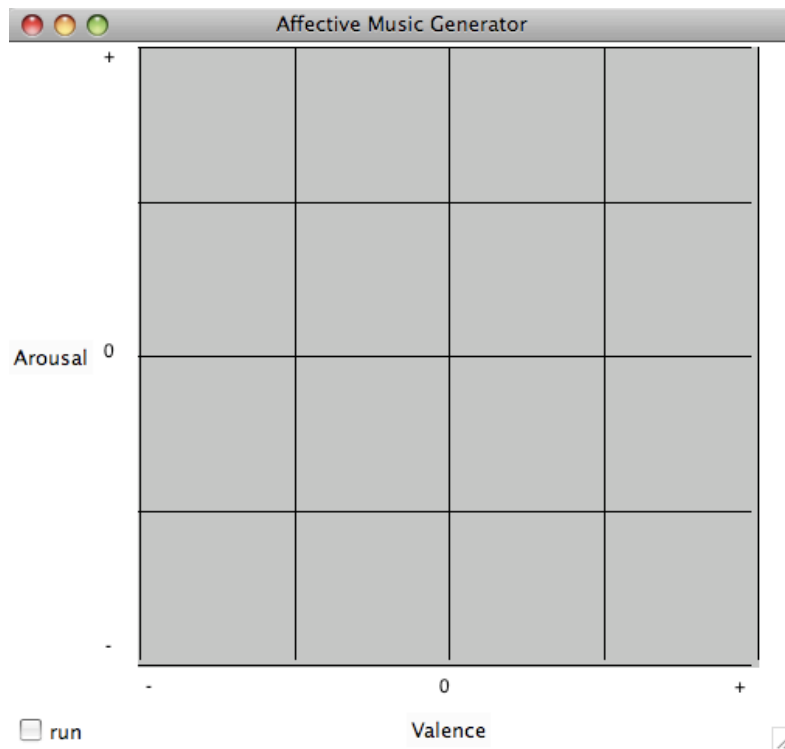
Figure 4.14: Clickable arousal-valence grid used in EMS study.

Each trial's clicked location was recorded along with the valence and arousal of the generated music. Results of correlations and t-tests between these datasets are given in table 4.5 (calculations by Janel Goodman). These show significant positive correlations between intended valence and perceived valence, and between intended arousal and perceived arousal. These also show no correlation between intended valence and perceived arousal. Ideally, there would also be no correlation between intended arousal and perceived valence, but there was a strong correlation. Perceptions of mood are highly complex, so this is not a surprising result. Figures 4.15 and 4.16 plot the mean perceived (i.e. clicked) valence or arousal for every

intended valence or arousal setting, and one standard deviation above and below. Data was not normalized before these operations. These results show that the music algorithm successfully generates music at the relative valence and arousal that it is set to; however, the valence and arousal parameters are not completely orthogonal, as evinced by the correlation between arousal settings and perceived valence.
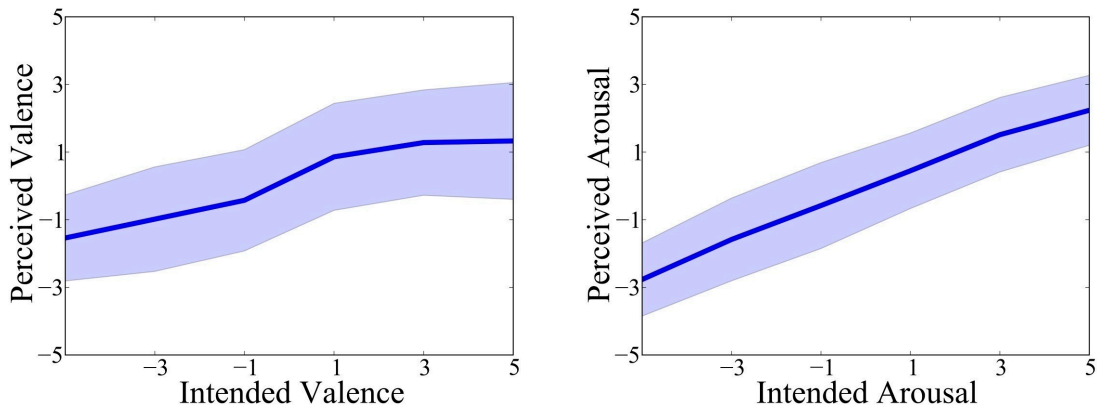
(a) Intended valence vs. perceived valence. (b) Intended arousal vs. perceived arousal.

Figure 4.15: Within-dimensions results of emotional music synthesis evaluation.
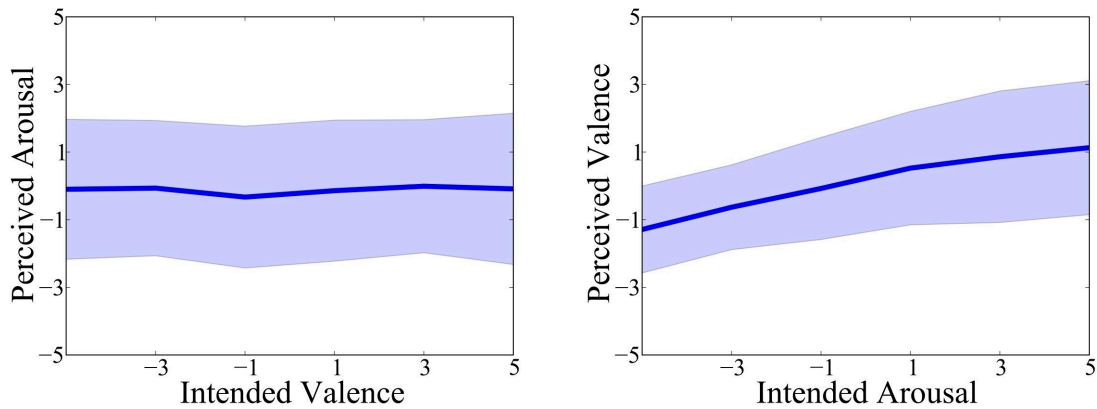
(a) Intended valence vs. perceived arousal. (b) Intended arousal vs. perceived valence.

Figure 4.16: Between-dimensions results of emotional music synthesis evaluation.

	Intended Valence	Intended Arousal
Perceived Valence	$r = 0.475$ $(p < .001)$, $t = 36.82$ $(p<.001)$	$r = 0.364$ $(p < .001)$, $t = 18.09$ $(p<.001)$
Perceived Arousal	$r = 0.008$ $(p < .001)$, $t = 0.44$ $(p<.66)$	$r = 0.679$ $(p < .001)$, $t = 23.61$ $(p<.001)$

Table 4.5: Correlations between perceived and intended arousal and valence.

4.2.2 Analysis

This system was not deployed in longer-term contexts; most users experienced it only once or twice during showcases or conference exhibitions. According to the framework, this system, if deployed in the right setting, should draw volitional usage and engagement over shorter terms. My recollections of user reactions support this conclusion. Even though most people are familiar with slideshows, many users seemed to feel that the concept was novel. ESG is simple to use; it has only one interface element. As will be discussed in greater depth later in this discussion, the generative content can make the system cognitively rich. Therefore, if the deployment setting provides promotion, convenience, and targeting—as is the case in showcases and exhibitions—shorter-term engagement will result. According to the framework, it seems unlikely that ESG would draw much volitional usage over longer terms. The design does not address cooperation, demonstrability, or ownership. The system is not immediate in longer-term contexts—a downloadable or web-based distribution, which might have addressed this, has not been created.

At first glance, the fact that there is only one menu-based interface element might seem to limit explorability and incrementality. Users who spend more time, however, may discover this is not the case. Many psychological studies suggest that humans naturally tend to find patterns in random sequences and make associations between random things (cf. Skinner, 1948; Whitson and Galinsky, 2008; Gilovich et al., 1985). The image sequence is related only by emotional keyword, therefore it is almost random; however, users can create relationships

between images and construct narratives out of the content. Users who do this introduce, for themselves, an activity—story-making—that contains explorability and incrementality. This activity also provides a sense of ownership, because users define the system's meaning to suit themselves (this makes ESG an example of an enchanting user interface as per Sengers et al., 2008). If they relate their stories to others, users also introduce demonstrability. Therefore, according to the framework, users who do these things should exhibit longer-term volitional usage than users who do not.

ESG engages users through the senses, and requires little in the way of direct user control; thus, it can be described as an immersive system, and users who become absorbed with ESG can be said to be immersed. In day-to-day life, immersive media (e.g. television, film, etc.) often uses pre-composed story-lines to hold the attention of viewers. Through unfolding plots, these story-lines continuously introduce new content to viewers, and through manipulating the mood of the content they cause viewers to feel emotionally invested. ESG, through compiling endless non-repetitive sequences of images, also continuously introduces new content to users, but the images do not belong to any pre-composed story-line. Thus, ESG poses a question: is the recipe for longer-term immersion and volitional usage simply: a) explorability (i.e. endless new content), and b) ownership (i.e. emotional investment)? Or are pre-composed story-lines necessary? It is hard to imagine ESG growing a user-base that is as large or as dedicated as the viewership of certain TV shows. There are two possible reasons for this: a) media containing pre-composed story-lines is somehow more immersive than media without, or b) design factors keep ESG from being as immersive as it could possibly be.

Let us explore the first of these reasons, by treating plot/story-line as a representational modality. Due to our natural empathic tendencies when given objects of emotional focus, story-line could be uniquely effective, in synchronization with mood, for eliciting emotional investment. If this is the case, users engaged in story-making could find ESG as immersive as television, and if they relate their stories to audiences, those audiences may also find the stories

139

immersive. Also, it is less costly, in terms of effort, to consume a story-line that is pre-composed than it is to construct one out of images. If this is a factor, users listening to the stories of ESG story-makers might actually be more immersed than the story-makers themselves (although the story-makers may also experience flow, and the listeners will not).

The first possible reason is not mutually exclusive with the second possible reason. Like AMRR, ESG was instrumental in my adoption of the channel model of modalities; especially the rule of concision, because—as in AMRR and ESG—violating the rule of concision provides support for the existence of the rule of concision, by inducing boredom or annoyance in users. Redundance within modalities exists in both systems, happening in different ways, and is problematic in each. ESG does not use excessive repetition like AMRR, but in ESG the mood and content is homogeneous over time, limiting the perceived explorability of the system. Mood is a modality; if the content presents the same mood for too long, it can lead to boredom. My recollections of user behavior support this statement, because users tended to move on to previously unexplored moods at frequent intervals, exploring many of them before stopping the session. However, even if there were unlimited moods to explore, the overall content might begin to run together over time. For example, even though the music does not repeat itself, it is quite self-similar. After enough exposure, it is possible that even changing musical mood will not make the music seem as though it has more to explore.

If I were to re-design ESG from scratch, I would design it for distribution via the internet for greater convenience. More importantly, however, I would explicitly address the story-making aspect of the system. Rather than giving users a drop-down menu to select moods from, I would let them draw mood arcs that, when played, progress over time. The emotion in stories is frequently thought of in terms of arcs, so this feature would emulate story-arcs while addressing the problem of mood sameness over time. I would allow users to upload their own music, attach time-stamped emotion tags, and generate slideshows according to the emotional arcs implicit in the music. This would offer users a greater level of ownership. Lastly, I would

integrate an audio-video recorder into the slideshow generator, so that users can improvise narrations or lyrics to slideshows and capture, then share, their creations. I could use this hypothetical new verion of ESG as a platform for experimentation on volitional usage and engagement. For example, I could run experiments looking at differences in participation between story watchers verses story-makers, in order to identify the impact of effort costs in practice-oriented systems. Or, by measuring how frequently people share their slideshows, I could further identify the impact of demonstrability.

I will finish this discussion with an aesthetic observation. Images are emotive; humans can perceive emotion and mood in them. Not only is this apparent in day-to-day life, the empirical validation of affective image databases such as the International Affective Picture System (IAPS) proves this (Lang et al., 2005). However, the technical challenges involved in the computational extraction of emotional meaning from images have, thus far, proved insurmountable. ESG is based on the following logic: if people upload images to Flickr with titles or keyword tags containing emotion words, then the images probably depict the emotion in some way. This is not always the case, and there is nothing that ensures the emotional accuracy of the slides. With that said, the generated slideshow is, in the overall sense, remarkably apt. Sometimes unexpected connections are made which serve to highlight the use of social media; for example, setting the interface to 'serene' might result in a slideshow consisting of mostly sleeping babies, sleeping pets, and landscapes—but once in a while, a bottle of pills might appear. In my view, these unexpected images are desirable, as ESG was created mainly as an artistic/aesthetic design artifact, not for any clinical use. If ESG needed to be emotionally accurate for scientific purposes, Flickr would need to be replaced with an image database, such as IAPS, with images that have scientifically-measured emotional qualities.

4.3 *Thrii*

The transitional project described in this section is smaller-scale than AMRR or ESG— it was created over the course of a month by a team consisting of Stjepan Rajko, Nicole Lehrer, Loren Olsen, David Tinapple, Assegid Kidane, David Lorig, Tatyana Koziupa, Diana Siwiak, Michael Baran, and myself. *Thrii* is an interactive art installation, designed for showcasing in demos and museum settings. It requires multiple users, and is designed to create an experience that feels ceremonial and communal. Aside from this goal, the *Thrii* design team had two design priorities. First, there was a high priority on visual and musical aesthetics so that audience members, as well as users, might find the system compelling. Second, there was a high priority on explorability, because we were intrigued with the idea that users exploring *Thrii* might begin to fall into ritualistic movement patterns. For this reason, the modeling component of *Thrii* is devoted to detecting repetitive or synchronized movements, and all feedback is mapped to this data. My role was music programmer, so I worked—with input from Diana Siwiak, Tatyana Koziupa, and others—to create the interactive music component. This system was described fully in Lehrer et al. (2010), which I am an author on, and some of the language in this section— namely musical description contributed by myself—is from that publication.

4.3.1 *Description*

As hinted by its name, *Thrii* has a geometrical design theme: there is three of everything in this system. Three users stand around a three-sided projection surface (called a 'half-cube' because it is shaped like a box with three sides removed, balanced on its corners), and each user holds a spherical glowing sensor (called 'glow-sphere'). Three speakers distributed equally about the space play localized music. As seen in the layout diagram given in fig. 4.17, *Thrii* is usually installed by itself in a stark, plain, darkened room. The only light comes from: a) the glow-spheres, b) ambient lighting from a lamp under the half-cube, and, c) a projection that is

142

mapped onto all three sides of the half-cube, coming from a ceiling mounted projector with a mirror redirecting the beam downward.
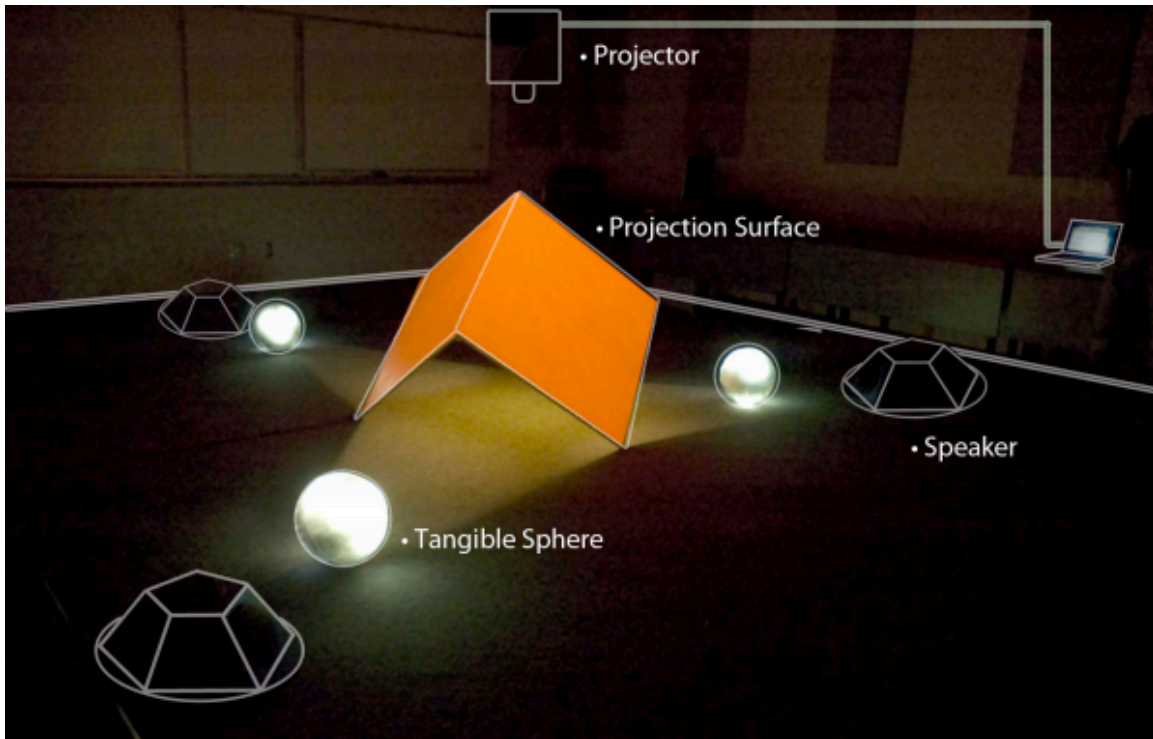
Figure 4.17: A diagram of the layout of *Thrii*. By Nicole Lehrer. Used with permission.

Each glow-sphere contains a Wii remote powered by a rechargeable battery. This affords two-way communication via bluetooth between each glow-sphere and the modeling component of *Thrii*. The circuitry of the remotes are modified to control the lighting of the spheres. Wii remotes have four small light-emitting diodes (LEDs) that can be individually switched on and off via blue-tooth, but in the glow-sphere remotes these circuits have each been re-routed to power a group of three larger LEDs. These groups of LEDs are what make the spheres glow; each glow-sphere contains twelve LEDs distributed equally so that each LED rests a few inches below the surface of the sphere and glows through diffusive material. Therefore, the system modeling component can turn lighting sectors of each glow-sphere on and off in patterned ways.

Wii remotes contain three-axis accelerometers, and data from these accelerometers is transmitted to the modeling component to be mapped to visual and musical feedback. The modeling component of *Thrii* is intertwined with the visual feedback component. Any movement of the glow-spheres causes patterns of pink and green light to cascade down the faces of the half-cube. The core of the modeling component is a dynamic time-warping (DTW) algorithm implemented by Stjepan Rajko. DTW is useful for calculating pairwise similarity, in real-time, between streams of data (cf. Myers and Habiner, 1981). Since there are three glow-spheres, three pairs of data-streams can be compared using DTW. One element of the algorithm involves windowing two data streams (with windows of 300 samples), then calculating the similarity between each sample of the first window with each sample of the second window. This results in a 300×300 similarity matrix. The similarity matrix for each pair of glow-spheres is visualized on each face of the half-cube, where the similarity values are mapped to pixels along a pink-to-green color palette. Green areas on a face of the half-cube correspond to areas of high-similarity in the DTW similarity matrix calculated between a pair of data-streams. Because each new data-frame from the glow-sphere accelerometers pushes an old data-frame out of the DTW window, these green areas cascade down the sides of the half-cube in complex patterns.

Calculating a Fast Fourier Transform (FFT) over each similarity matrix shows the frequency over time at which any movements are repeated. This affords exploration of time-delayed movement emulations by users. The FFT is also used to find a level of repetitiveness, and frequency of repetition, within the movement of single glow-spheres. In addition to the FFT, the DTW involves a process borrowed from the Levenshtein Distance algorithm to find a path of maximum similarity traversing each similarity matrix (cf. Levenshtein, 1966), by which an overall running similarity value can be determined for each pair of data-streams. The running similarity and repetitiveness values are mapped to various forms of feedback. For example, the running similarity values between glow-spheres control the internal lighting patterns of the

Interaction	Feedback
Move any glow-sphere	Transition to episode 2

Table 4.6: *Thrii*, musical episode 1, feedback mappings.

Interaction	Feedback
Glow-sphere 1 repetitiveness	Drone amplitude LFO rate
Highest similarity pair	Arpeggiator volume
Glow-sphere 3 repetitiveness	Arpeggiator speed
Three-way similarity	Clarinet mode
Similarity: glow-spheres 1 & 2	Clarinet breath pressure
Glow-sphere 3 repetitiveness	Clarinet reed pressure

Table 4.7: *Thrii*, musical episode 2, feedback mappings.

Interaction	Feedback
Three-way similarity	Triggers power-up
Overall repetitiveness	Drum volume
Glow-sphere 1 repetitiveness	FM synth volume
Glow-sphere 2 repetitiveness	FM modulation width
Glow-sphere 3 repetitiveness	FM modulation rate

Table 4.8: *Thrii*, musical episode 3, feedback mappings.

glow-spheres. If two glow-spheres move in synchrony, LED sectors within the glow-spheres will also blink in synchrony.

Running similarity and repetitiveness are used to drive musical elements as well, and these mappings are given in tables 4.6 to 4.8. *Thrii* contains three interactive musical episodes, described in the rest of this subsection. If the glow-spheres are left alone for three minutes, the music transitions to episode 1, and stays there until someone moves a glow-sphere (as shown in table 4.6). Thus, episode 1 is a passive listening environment, designed to surreptitiously

fade into the background for the benefit of any non-users in the area. Otherwise, if *Thrii* were installed in an area with employees stationed nearby, those employees might become annoyed with the music. Episode 1 is composed to evoke a ghostly quality, and sounds reminiscent of large objects moving and creaking underwater.

Once participants pick up glow-spheres to begin causing green lights to play down the sides of the half-cube, the music switches into episode 2. The stated theme of episode 2 is synchrony, so its mappings are designed, in part, to elicit synchronous movements between the users. It is orchestrated with three instruments (see table 4.7). First, there is a continuous drone with the amplitude modulated by a low-frequency oscillator (LFO). The frequency of the LFO is mapped to repetitiveness in glow-sphere 1, so doing repetitive movements with that glow-sphere increases the modulation rate. Second, there is an arpeggiator that plays note patterns. The speed at which the arpeggiator cycles through notes is mapped to glow-sphere 3, so that more repetitive movement increases the rate of the arpeggiator. Also, the arpeggiator fades in and out in volume, mapped to the running similarity of whichever glow-sphere pair is currently most similar. Lastly, there is a sound, generated by a physical model of a clarinet, playing stochastically-generated notes within a diatonic mode. The overall similarity of movement between all three users (i.e. the mathematical product of the three running similarity values) determines the harmonic mode the clarinet plays in: low similarity is mapped to a minor mode, and high similarity is mapped to a major mode. The clarinet physical model takes a breath pressure parameter that controls volume, and this is mapped to the running similarity between glow-spheres 1 and 2 such that high similarity means high breath pressure. Also, the physical model takes a reed pressure parameter, combining pitch modulation and amplitude modulation in a way that sounds clarinet-like. This parameter is mapped to repetitiveness in glow-sphere 3.

Episode 2 lasts three minutes, after which episode 3 begins. The stated theme of episode 3 is repetition, so its mappings are designed, in part, to elicit repetitive movement

from each user. Episode 3 is orchestrated with two interactive instruments played over a non-interactive bed of windy sounds reminiscent of blown bottles. The first interactive instrument is a frequency-modulation (FM) synthesizer with carrier centered on middle C. The user holding glow-sphere 2 controls the FM modulation rate, with more repetitive movement mapped to higher modulation rates. The user holding glow-sphere 3 controls the FM modulation depth, with more repetitive movement mapped to deeper modulation. The user holding glow-sphere 1 controls the overall volume of this instrument through repetitive movement, and the mapping is such that the instrument is inaudible when there is no repetitive movement, and loud when there is highly repetitive movement. The second interactive instrument is a drum pattern that fades in if all users perform repetitive movement simultaneously. The users need not be moving synchronously and repetitively in order to bring in the drum pattern, but if they do move synchronously, a short ascending sound, called a 'power-up' due to its resemblance to the power-up sounds in certain old video games, will trigger if users perform highly synchronous movement. Episode 3 lasts three minutes, after which the music will transition back to episode 2.

4.3.2   Analysis

Thrii was designed mainly for installation at showcases and museum environments. According to the framework, *Thrii* should draw volitional usage in the shorter term given these settings, and my recollection of user responses to the system supports this. For example, in the showcases where *Thrii* was exhibited, where many interactive installations were presented, it was often the most well-attended piece. Crowds frequently formed around the half-cube, and the glow-spheres would be passed around among members of the audience who wanted to try the system. *Thrii* is novel and cognitively rich. It has many usability problems—more on this later—but in showcase settings, members of the design team can help users overcome these. The showcase settings also provide promotion, convenience, and targeting.

Figure 4.18: *Thrii* in use during a showcase.

On the other hand, according to the framework, *Thrii* does not seem as likely to elicit longer-term volitional usage. *Thrii* is explorable, demonstrative, and cooperative, but there are problematic aspects with regard to the other longer-term heuristics. Immediacy is poor, for example: the convenience *Thrii* has in showcase settings does not extend elsewhere, because it takes several people to set the system up and run it, and the usability of the system is low, because the affordances of the system are not easily perceived by novices. The three glow-spheres each have distinct relationships to the music and to each other, but they look identical so there is no easy way to tell which glow-sphere is which. For example, looking at table 4.7, we see that in episode 2, the movement repetitiveness of glow-sphere 1 is mapped to the speed of an LFO. The only way a user can tell whether they have glow-sphere 1 or not, however, is through interacting with it and listening to the results. Because the sonifications are difficult

for novices to parse, this creates a cyclical problem: for users, knowing which glow-sphere one is holding is difficult because it is hard to understand the sonifications, and it is even harder to understand the sonifications if the user does not know which glow-sphere is being held.

There is a similar problem with ambiguity between the glow-spheres and the faces of the half-cube. Each glow-sphere impacts two, and only two, of the faces of the half-cube, because it is paired with the other glow-spheres to generate two of the three similarity matrixes. This implies a prime standing location for each glow-sphere holder near a specific corner of the half-cube, allowing the user to see both impacted faces. However, just as the glow-spheres are visually identical, the corners of the half-cube are similar in appearance, so users sometimes have a difficult time knowing where to stand to observe their own interactions.

For the sake of musical aesthetics, many simultaneous sonification mappings were implemented in episodes 2 and 3. This allows users to generate subtle and complex music, but also violates the the channel model rule of coherence (as discussed in section 2.3), and this makes the sonifications more difficult to parse. With practice, listeners can begin to attain skill at parsing complex sonifications. As discussed in section 4.1.3, this skill attainment is easier if the sonifications are consistent over time and consistent within some rational structure. Neither of these things is true of sonifications in *Thrii*. The musical episodes, which were incorporated to add musical variety, use different mappings; therefore sonifications are not consistent over time. Also, the themes of these episodes are not thoroughly adhered to, because musical aesthetics occasionally called for a mapping to repetition being implemented in episode 2, or a mapping to synchrony being implemented in episode 3. Therefore, sonifications are not really consistent within any rational structure.

One other thing makes the sonifications, as well as other interactive elements such as the glow-sphere blinking, difficult to understand. The DTW algorithm calculating the running similarity values uses a window of 300 frames, covering a time-span of 6 seconds. The FFT calculating repetitiveness uses a similar window size. This means that the running calculations

are quite slow-moving. If two users suddenly go from exhibiting no movement synchrony to exhibiting perfect movement synchrony, they must continue moving in tandem for six seconds before the running similarity will have increased from zero to one. The same is true for a person that goes from exhibiting no movement repetition to high levels of movement repetition. The movement synchrony and movement repetition values drive much of the interaction, consequently, slowness and latency within them makes it harder for users to associate movement with resulting feedback.

If I were to redesign *Thrii* from scratch to draw longer-term volitional usage, I would do several things differently. In order to increase convenience, I would relax the restriction that three users are necessary for system operation. Each glow-sphere would be differently-colored, and each would produce unique music by itself in addition to producing music in conjunction with each of the other glow-spheres. This change would allow individual users to practice during times when others can not. It would also afford more ownership, as users could begin to specialize in one of the glow-spheres. In order to keep the music interesting, I would continue to use multiple sonification mappings despite the fact that this violates the rule of coherence; however, I would focus on ensuring that sonification mappings were learnable with practice. Instead of automatically-switching musical episodes, I would integrate manually-selected sonification environments. Then, users could practice environments they find appealing, and switch if they become boring. Within these environments, mappings would remain consistent for learnability. All sonification environments would contain one highly audible key mapping directing users where best to stand around the half-cube. For example, I might map volume to location so that users cannot hear their sonifications at all unless they are positioned near the corner of the half-cube affording the best view of their visual impacts. I might also design mappings to reward slow movement over fast movement, because slow movement is easier to synchronize among groups, and is conducive with the ceremonial feel of *Thrii*. As the half-cube and projection space are not practical for home installation, I would set *Thrii* up as a long-term

exhibit in a museum, gallery, arcade, or some other location where interested people could use the system whenever they wished.

The main reason *Thrii* was discussed here in this chapter, despite its relatively small scope compared with AMRR and ESG, is because it explores the concepts of demonstrability and cooperation more explicitly than other systems I have worked on. Three users are needed for full operation of the system, and many mappings require coordination between users, therefore, exploring the system to its fullest extent can only be done cooperatively. The half-cube, as a visual display that users stand around, is also designed around group dynamics; users position themselves to see both the visuals and each other. If these users begin emulating each other and using each other's movements as fodder for improvisation, they might begin to experience shared feelings of flow and euphoria just as ensembles of improvisatory musicians or actors do (cf. Sawyer, 2006). The aesthetics and performative nature of the system tend to act as a form of promotion and draw an audience; in most showcases the users are working in an environment where they perform, not just for each other, but for the crowd of onlookers as well.

This subsection, I have been critical of *Thrii*, especially with regard to its legibility. These legibility issues are why, in exhibitions of the system, cues from team members were required before desired movement patterns would emerge. Without these cues, users often drew mistaken conclusions about feedback mappings, and often guessed wrong about the intended function of the system. This points to a possible upside to the legibility problems, however: that *Thrii* might be classified as an example of an enchanting user interface as described in Sengers et al. (2008). Users can observe that they are changing things about the system, but they cannot discern exactly what they are changing; consequently, they are forced to assign meaning for themselves. Viewed in this way, the afforded activity can be described as "the activity of trying to identify interactions that control the system." This is a highly exploratory activity, requiring a great deal of concentration on the music and visuals and a great deal of

inter-user cooperation. Cognitive richness and, to a lesser extent, incrementality, emerge from this activity.

Chapter 5

THE FRAMEWORK AS DESIGN TOOL

In order for the framework to be an effective tool for designers, it must be useful in the analysis of systems (as discussed in chapter 4), and it must also be useful in the design of new systems. This chapter begins the process of validating that it satisfies this second requirement. A new experiential system called *pdMusic* is introduced, which I have designed and developed for the past two years. Here I will describe and analyze two iterations of this system in the same manner as I described and analyzed the systems in chapter 4. The common goals of chapter 4 systems (drawing usage and engagement, being aesthetically compelling, and informing the theory and praxis of experiential systems design) are also goals of this system and will be discussed here. However, unlike those systems, work on *pdMusic* started after the work on the framework was initiated, and *pdMusic* is still in active iterative development. This has multiple implications. First, it affords the study of volitional usage on this system in a more formal way than was possible in those other systems. Second, because the design and development of *pdMusic* was guided by the design framework, we can begin to validate the idea that using the framework helps designers make systems which elicit volitional usage. Third, studying this system and its usage allows assessment of design framework criteria that could not otherwise be sufficiently addressed, such as the fit and modifiability of the framework.

As will be seen in upcoming discussions, performing research on the topic of volitional usage is fraught with challenge, and the reason for this is simple: small details, seemingly of little consequence, often act as inhibitors to volitional usage and engagement. Therefore, any contrivance forced upon users in the name of research will depress their volitional usage. The more a line of research attempts to quantify exact levels of engagement and volitional usage, the more these detrimental contrivances seem to be required. For this reason, at the time of

this writing I have settled on a guiding tenet for my research protocols which can be summed up in one word: naturalness. Studies of system usage should be performed in the most natural environments possible, with minimal hassles or research contrivances impeding volitional usage. This naturalist approach limits the types of data collection which can be done. For example, it rules out the use of things like physiological sensing, formal interviews, or video recording—all of these would change the way people behave. This approach lends itself more to qualitative research practices than quantitative research practices, and the research I do in this chapter, though ultimately mixed methods research, depends more on qualitative research methods than quantitative. All quantitative data here results from quantitative methods embedded within qualitative research, and is designed to support or fine-tune smaller portions of the design framework. Per Creswell (2008), some of the distinguishing characteristics of qualitative research are: a) research in natural settings, b) the use of a theoretical lens, and c) an emergent research design.

My work shares these three characteristics, and also follows another pattern of qualitative research described by Creswell: "In a qualitative study, inquirers state research questions, not objectives (i.e., specific goals for research) or hypotheses (i.e., predictions that involve variables and statistical tests)" (pg. 129). Formal hypotheses would be difficult to construct in the context of my large multifaceted framework. For example, any statement such as "I hypothesize that designing systems using my framework results in systems drawing more volitional usage than other systems which were not designed using my framework" can not be convincingly proven or disproven without testing a large number of distinct systems that were created using my framework against other systems that were designed without my framework. Such an experiment would not be practical, so my research here is based on research questions rather than hypotheses. My central research question, and sub-questions therein, are:

- Given the opportunity to use *pdMusic*, how do people respond?

- How do users and audience observers describe the experience?

- In seeking to understand volitional usage, what issues arise?

- What emergent theories explain the behavior of users, and are these theories in line with the design framework?

The two iterations of *pdMusic* discussed in this chapter were designed in the effort of gaining insight into the research questions above. Per Zimmerman et al. (2007), *pdMusic* fits the definition of a research artifact, and that places this work squarely in the realm of research through design. As will be seen, the first design iteration was designed for the purpose of exploring longer-term heuristics, but the process of design led me to the realization that shorter-term heuristics needed to be explored first. For this reason, the data analysis in section 5.3 does not provide support for all heuristics of my design framework. The ultimate goal of this research is not to prove hypotheses, it is to generate them. The framework as given in chapter 3 is logically consistent and has a strong element of common sense, and for this reason I believe it has value as a collection of loose design guidelines. Empirical validation of some of these ideas, however, will come with future testing.

5.1   First Iteration of *pdMusic*

This section describes the first iteration of *pdMusic*, which I created with guidance from Prof. Todd Ingalls. When this project was first conceptualized its main area of application was intended to be Parkinson's Disease rehabilitation; *pdMusic* was intended to be a collaboration with Catherine Vuong, and was viewed as an extension on the system she describes in Vuong et al. (2011). Research suggests early-stage Parkinson's disease patients can slow deterioration in some symptoms by performing wide-amplitude movement exercises as part of their physical therapy (Farley and Koshland, 2005). Therefore, a practice-oriented system such as *pdMusic* could be designed to encourage wide-amplitude movement. Scheduling concerns precluded this collaboration, however. Although *pdMusic* may still prove useful for movement rehabilitation, its main function has since been repurposed: now it is designed to serve as a case study on volitional usage and to afford experimentation and research on the topic. The concept for this iteration of *pdMusic* is briefly discussed in Wallis et al. (2013).

pdMusic affords three interrelated activities: it is a musical instrument, it is a rhythm game, and it is a platform for composition and musical training. Rhythm games, as exemplified by popular video games such as *Wii Fit*[1], *Guitar Hero*[2], and *Dance Dance Revolution*[3], challenge players to closely synchronize with a complex sequence of actions. This sequence of actions is given by the game, and action timing is usually related to a background song in some way. For this reason, it is useful to think of each sequence-song combination as a singular game session; most commercial rhythm games contain large libraries of these game sessions. Existing rhythm games are fairly well designed in terms of the heuristics related to mastery. For example, they have incrementality, because their libraries of game sessions are large and diverse enough that players at all levels will find sessions of an appropriate difficulty. Commercially successful games

---

[1] see http://wiifit.com/

[2] see http://hub.guitarhero.com/

[3] see http://www.konami.com/ddr

also have a measure of immediacy, as they are relatively low-cost and easy to set up. Rhythm games are somewhat limited in potential complexity: once the most difficult game sessions can be defeated with ease, players are unlikely to continue being engaged. As a convenient short-hand, in this chapter *pdMusic* game sessions will be called 'compositions.'

5.1.1   Description
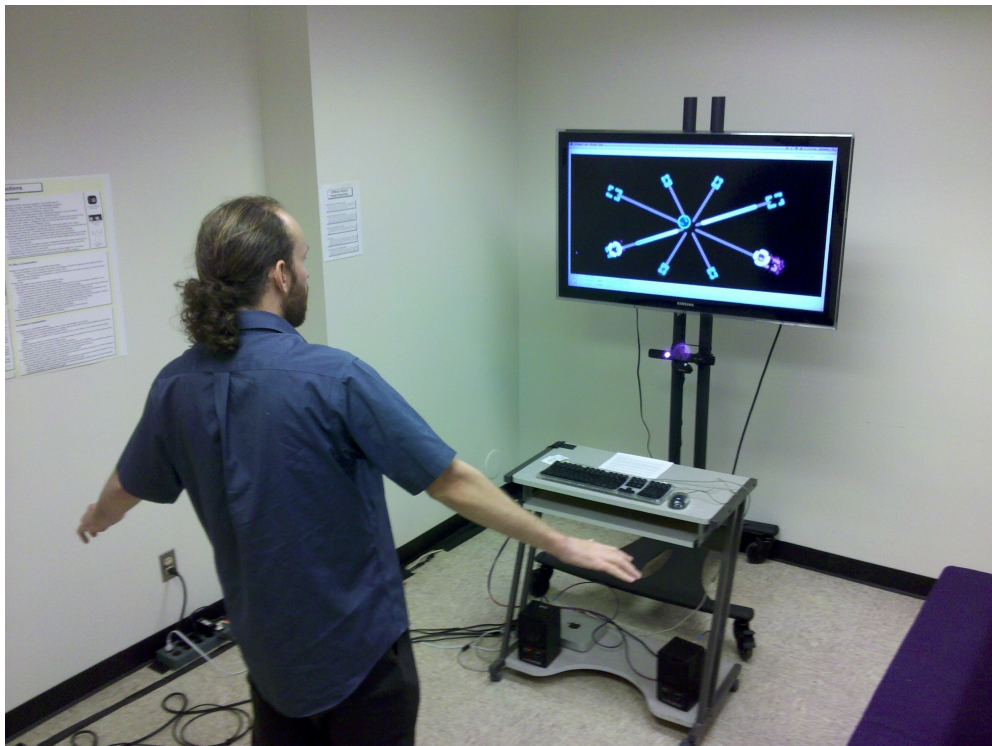
Figure 5.1: The author demonstrating system usage.

As seen in fig. 5.1, visual feedback in *pdMusic* is presented through a large monitor placed approximately at eye level for a standing user, and musical feedback is presented through a pair of high-quality computer speakers. Configuration of *pdMusic* is done using a mouse and keyboard set at waist level. Sensing in *pdMusic* is based on optical motion capture via a Microsoft Kinect sensor. When participants face the video monitor, the Kinect, which is placed directly below the monitor and aimed back at the user, senses their movements. This sensor creates a depth map consisting of everything in its field of view and, by tracking three-dimensional

location of many of the user's joints, allows skeletal models of each user to be created. When users enter the system, they are first required to stand in a specific pose, called the 'psi pose' (alternatively called the 'surrender pose', due to its resemblance to a similar pose many film characters make when threatened with a gun—see fig. 5.2). The Kinect recognizes this pose, and uses it to construct and calibrate an internal model of the user's skeleton.

Figure 5.2: Psi pose required for Kinect calibration

Once a user's skeleton is recognized, visual elements begin to track his or her movements. The case study system uses the visual interface shown in fig. 5.3. The blue orb near the center is called the 'center-of-motion indicator,' and tracks the two-dimensional location of the user's throat within the Kinect field-of-view. The location of the user's throat does not factor into any other feedback; the reason the center-of-motion indicator tracks the throat is simply to help users situate themselves. That is, the throat is the skeletal center-point on the axis create by the arms, so freedom of arm movement will be maximized if users stand in a location where the throat is central to the Kinect field-of-view. The two white orbs in fig. 5.3 are called 'hand cursors,' tracking the two-dimensional location of the user's hands within the Kinect field-of-view. These are the main interface modalities through which users control *pdMusic*.

The green rectangles distributed in an ellipse around the *pdMusic* interface are called 'targets.' These do not move, but they interact with the hand cursors to generate music. Whenever a hand cursor touches a target, a musical note plays. These notes are generated by a hosted Au-
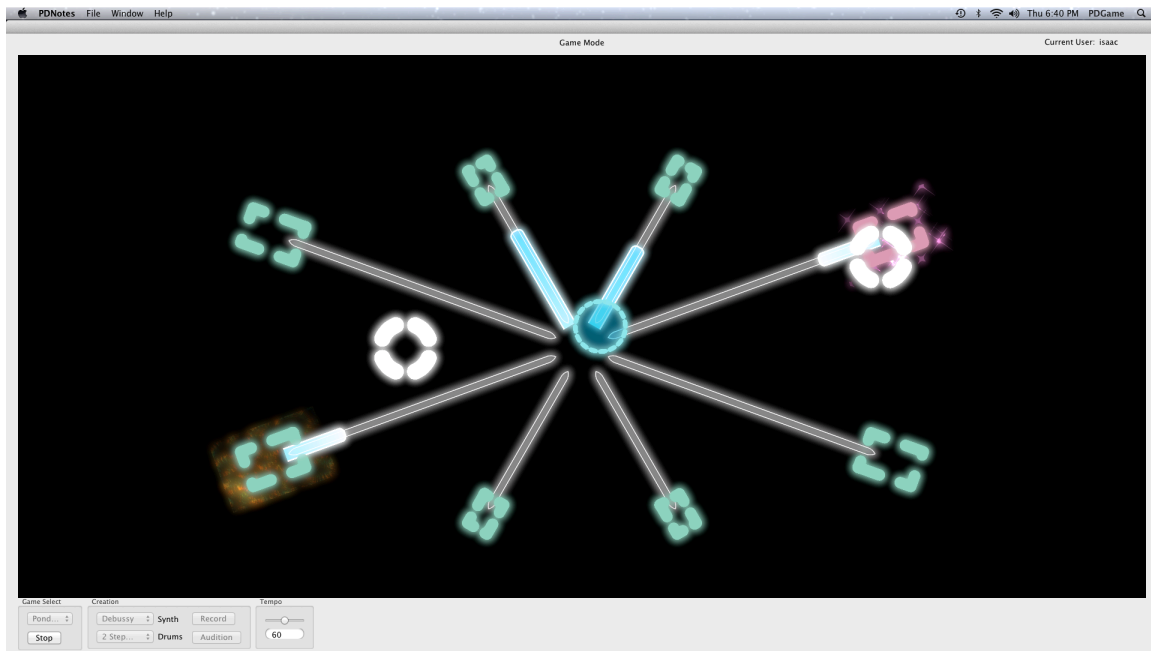
Figure 5.3: First iteration of *pdMusic* interface.

dioUnit plugin, Green Oak's *Crystal*[4]. The timbral setting of *Crystal* is configured with a drop-down menu on *pdMusic*'s interface. This menu contains a list of preset timbres, or if 'Configure Timbre' is selected, *Crystal*'s interface appears. *Crystal* is a highly flexible synthesis plugin, affording *pdMusic* an effectively infinite number of timbral variations. Another drop-down menu allows users to enable and select a percussive rhythm pattern to use as a metronome. The beats are generated by another hosted AudioUnit plugin, *iDrum*[5]. The drop-down menu contains a list of preset timbres, and if the 'Configure Beat' option is selected, the *iDrum* interface will appear. *iDrum* is a highly flexible rhythm sequencer, so an effectively infinite number of rhythms can be composed. Metronomes are useful tools for beginning musicians on any instrument; they help people keep in time when playing rhythmically. They also give people a rhythmic framework which can be used as fodder for improvisation. The different beats lend themselves to different styles of play. A slider controls the tempo (in bpm) of the rhythmic patterns, and this slider also controls the tempo of any rhythmic elements in the note synthesis.

---

[4]see http://www.greenoak.com/crystal/Crystal/Crystal.html

[5]see http://www.izotope.com/products/audio/idrum/

My first priority as a designer was to ensure that *pdMusic* affords playing and improvising music; all other functionality is built upon this afforded activity. In the pursuit of this, one challenge I encountered was the tyranny-of-choice involved with choosing a pitch layout; i.e., choosing the musical note to be played by each of the eight targets. Numerous pitch layouts could have been implemented, each with distinct benefits and drawbacks. In the end, the approach taken to solving this problem involved two phases. First, a set of notes was selected through researching the scales and modes used in existing music from appropriate genres. Since *pdMusic* incorporates a full-featured synthesizer and a drum sequencer, it is capable of emulating electronic dance music, such as the genres and sub-genres of house music. Therefore, I obtained a corpus of 59 songs, chosen by a local electronic music DJ (Andrea Silkey, who DJs under the name Sonique des Fleurs), and ascertained the key of each song. The list of songs analyzed, and the keys therein, can be seen in appendix A. The overwhelming majority were in a minor key. For this reason, I selected a set of notes containing one octave of a natural minor (i.e. aeolian) mode. The natural minor was chosen because, unlike other minor scales such as the harmonic or melodic minor, it is one of the seven diatonic modes discussed in section 4.2.1. This affords a greater degree of flexibility; users can switch between modes—for example, to the major (i.e. ionian) mode—simply by changing the style-of-play to emphasize or deemphasize certain notes within the set.

Once a set of notes was selected, the second part of the pitch layout design problem was the placement of each note on the ring of targets. The approach to this problem was solved in part through looking at the pitch layouts of existing musical instruments. *pdMusic* is a pitched instrument that affords polyphony; therefore, it allows musicians to play both melodies and harmonies. This poses a design challenge because a layout that is optimized for melodic playing differs from a layout that is optimized for harmonic playing. Statistically, melodies consist predominantly of step-wise motion, or motion based on 2nd intervals—intervallic leaps do occur in most melodies, but take place with less frequency (Schellenberg, 1996). Harmonies,
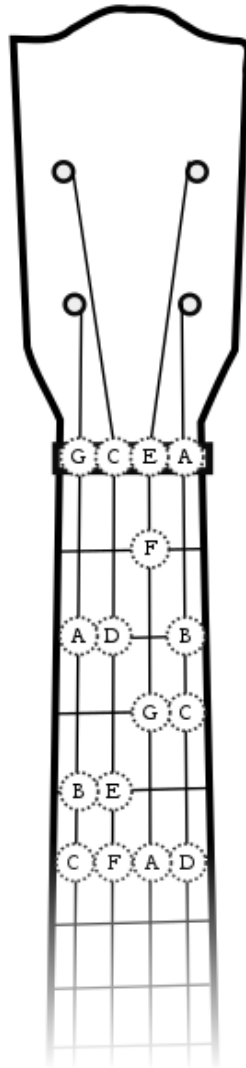
Figure 5.4: Pitch layout of the first five frets on a ukulele.
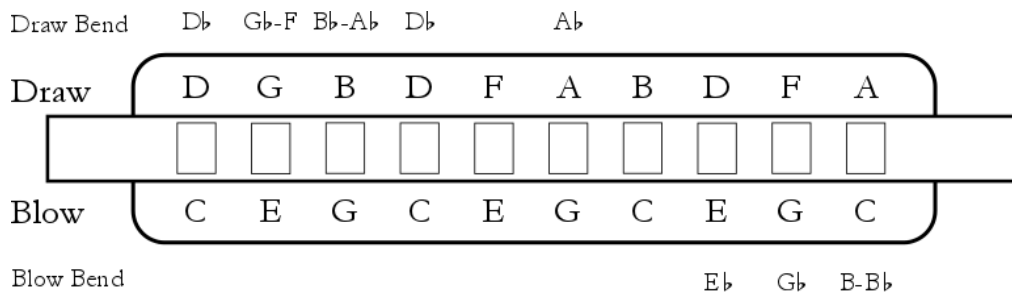
Figure 5.5: The pitch layout of a diatonic harmonica in the key of C major.

on the other hand, usually consist of triadic intervals (i.e. 3rd or 4th intervals). Therefore, a layout optimized for melodic playing would facilitate stepwise motion, whereas a layout optimized for harmonic playing would facilitate triadic motion. Let us consider two instruments that I know well: ukulele and harmonica. These instruments are not similar in form; the only thing they have in common is the ability to play both melodically and harmonically. In each of these, the two playing styles are balanced by mapping stepwise motion and triadic motion to two orthogonal playing techniques that are afforded by the instrument. In ukulele, stepwise notes are laid along the length of a neck on a single string, and triadic leaps are mapped across the neck on different strings (see fig. 5.4). In harmonica, near step-wise notes are achieved by alternating direction of breath, and near-triadic notes are laid out along the length of the instrument if breath direction is not changed (see fig. 5.5).

Because *pdMusic* users stand in the center of a ring of targets and play them by touching them with their hands, two obvious playing techniques emerge from the affordances of this design. The first, which I call 'arm sweeping,' is when users perform radial movement with one arm, moving from target to target. This playing technique facilitates movement between adjacent targets on the ring. The second, which I call 'alternating bilateralism,' is when users play using alternating hands. This playing technique facilitates movement between targets on opposite sides of the interface. I viewed this same-hand versus alternating-hand framework as analogous to breath direction in harmonica. For this reason, I used the note-mapping of harmonica as a model: I placed triadic intervals adjacently on the ring, and placed step-wise intervals oppositely on the ring. Therefore, arm sweeping produces chordal arpeggiations, and alternating bilateralism produces scalar motion. This resulted in a layout of a ring of pitches that were defined in relationship to one another, but does not imply any specific rotation of this ring of pitches.

There is another playing technique afforded by the design of *pdMusic*, which I call 'pair-wise bilateralism.' In this, users play two targets simultaneously. This can be done in an axial

way, where the arms are fixed in relation to one another and rotated as if on a fulcrum, or in a wing-like way, where the arms are raised or lowered together. Through this playing technique, users can play two-note chords and harmonic progressions. In harmonica, chords can also be played through widening the embouchure. Because harmonica players are limited to two breath directions, however, harmonicas were designed to facilitate certain commonly-needed chords over others. On standard diatonic harmonica, the notes are mapped such that outward breath produces notes belonging to the I chord, which is the tonic chord in the manufacturer-specified key of the harmonica, and inward breath produces notes belonging to the V^9, which is the dominant chord (with upper extensions) in the key of the harmonica. The tonic and dominant chords are the two most important chords in many styles of music, so this pitch layout, by facilitating the playing of these chords and inhibiting the playing of out-of-key notes, makes harmonica easy for beginners. Therefore, I modeled the pitch layout of *pdMusic* on this, by mapping the rotation of the pitch-ring so that the notes of a tonic chord (i.e. the I chord) are on the top half of the ring, and the notes of a dominant chord (i.e. the VII^7 chord, in the natural minor) are on the bottom half of the ring. This results in the pitch layout shown in fig. 5.6. This layout facilitates the voicing of these two chords through wing-like pairwise bilateralism, and all other diatonic chords can also be played through some other sort of pairwise bilateralism. The only diatonic chord requiring both hands on the same side of the target ring is the VI chord (i.e. A♭ major in the key of C minor).

My second priority as a designer was to ensure that *pdMusic* affords the creation of sharable skill-based products in order to foster ownership, incrementality, and demonstrability. Therefore, I integrated a composition platform that is combined with a rhythm game. For clarity, I will describe the rhythm game functionality first. The *pdMusic* interface contains a 'play/ stop' toggle button and a drop-down menu containing a list of available compositions. Each of these compositions contain a preset configuration—a snapshot—of the timbral, rhythmic, and tempo settings of the system, and selecting a composition from the menu causes *pdMusic*
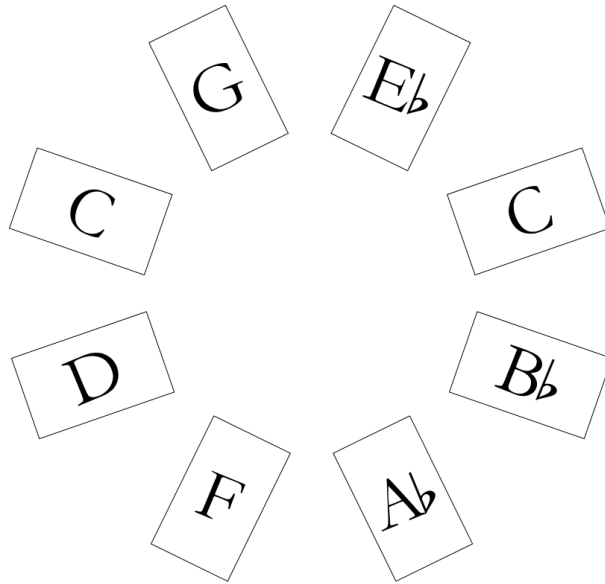
Figure 5.6: Pitch layout of *pdMusic* with targets and the attached notes.

to be configured with the composition settings. If desired, a user can change the configuration before playing the composition.

Once a user presses play, the configured rhythm begins to play and 'note indicators' begin to appear. As seen in fig. 5.3, the interface contains lines, called 'note-paths,' radiating from the center out to the targets. The two note paths in the top center have note indicators traveling along them, where the note indicators appear as wide lines, each containing a color gradient of blue fading to white along their length. Note indicators move outward along the length of the note-paths, disappearing from view as they pass into their respective targets. Whenever a note indicator intersects with a target, one of three audio-visual indications will take place, depending on the actions of the user. If a hand cursor is held over a target while a note indicator is passing into that target, pink sparkles will appear behind the target and the note attached to that target will be played in the currently-configured timbre as usual—the sparkles indicate that the note is being played correctly. If no hand cursor is touching a target, but a note indicator is, a yellowish-green particle effect will appear behind the target, and a scratchy noise will sound—this is designed to alert the user to the fact a note is being missed. Or, if a

164

user plays a target when there is no note indicator touching it, the musical note attached to that target will still play, but the target will vibrate back and forth and a scratchy sound will play—this is designed to alert the user that an unnecessary note is being played. The pink sparkles are meant to evoke success or celebration, and the yellow-green particles are meant to evoke disappointment. Thus, these visual elements are designed to encourage users to touch targets concurrently with note indicators. If all the note indicators of a composition are successfully played in this manner, the music that is generated by the targets will be identical to a previously-recorded composition. In the first iteration of *pdMusic*, after each composition is attempted, the percentage of the composition's duration that was correctly played by the user is calculated and displayed for the user as a score.

Once users have attained some skill at playing and improvising within *pdMusic*, they can record their own compositions and share them with other users. The interface contains a 'record/stop' toggle button and an 'audition' button. When recording is enabled, the targets played and timings therein are saved in a file along with the currently configured tempo, timbral, and rhythmic settings. This file contains all the information necessary to construct a *pdMusic* composition. Once a composition is recorded, if the user wishes to save it permanently, the audition button must be pressed. The recorded composition will then play as if it had been selected and played from the composition drop-down menu. After the user plays through the audition game, a dialog box will appear to ask the user if the composition was deemed satisfactory to play; if so, the user can name and save the composition, and its name will appear in the drop-down menu of compositions available to all users.

5.1.2   *Comparison with Similar Activities*

The activity afforded by *pdMusic* has similarities with playing rhythm games and playing musical instruments. In the following paragraphs I provide a brief analysis of *pdMusic* by

using my design framework as an analytical tool and comparing the *pdMusic* activity to these previously-existing activities.

Rhythm Games: According to the framework, those who try *pdMusic* ought to persist within usage at rates that are comparable to the user persistence of many commercial rhythm games. The game sessions of commercial rhythm games constitute complex and challenging activities, but the object of synchronizing movement with action sequences is not very conducive to explorability—these games tell players precisely what to do and when to do it. Part of the logic underlying the design of *pdMusic* is that allowing users to improvise and create their own compositions is a more explorable paradigm. Volitional usage should be more open-ended because, unlike in commercial rhythm games where users are unlikely to continue playing once the toughest game sessions have been mastered, *pdMusic* users can create playable compositions that are even harder; this results in an ever-expanding library of playable compositions for all users. The heuristics of ownership and demonstrability should be improved, since composing music and choreographing movement are self-expressive and skill-expressive activities—sharing compositions is equivalent to showing off compositions. Last but not least, users can learn musical techniques by playing each other's compositions, thus improving system incrementality. This could also result in a more rapid pace of musical innovation among the *pdMusic* user base.

Musical Instruments: Since playing music with *pdMusic* does not require physically exciting a sound-generator, there is no haptic feedback. Haptic feedback is helpful in instrument learning, and this is part of the reason why other touch-less instruments, such as theremin, are considered difficult. For example, the sense of touch helps wind instrument players understand: a) how much breath pressure is applied, b) how the instrument is positioned, and c) what note is being played. As this sensory modality is unavailable to *pdMusic* users, achieving virtuosity on the instrument may require more effort.

166

Compared to acoustic or analog instruments, *pdMusic* contains more latency. Most digital instrumentation contains a small but perceptible amount of latency. In acoustic or analog instrumentation, exciting the musical interface automatically generates the sonic response because both things occur in one action. For example, plucking a ukulele string is equivalent to generating a ukulele note. In digital instrumentation, exciting the interface and generating the tone are completely separable; this affords the interchanging of interface methods (e.g. MIDI keyboards verses MIDI guitars) or tone generators (e.g. string timbres verses vocal synthesis). It also introduces a small but often perceptible amount of latency. In *pdMusic*, this latency is visible; the hand cursors often trail the physical movements of the hands by a small amount. This latency seems to be built into the Kinect sensor, and therefore cannot be eliminated without drastically redesigning the system. It also, as will be seen in section 5.2.1, decreases the usability of the interface; users perceive *pdMusic* as less controllable because of it.

This latency can be obscured through selecting timbres of slower attack, but this comes at the cost of perceived responsiveness. Each timbre lends itself to a different style of play. Timbres that fade slowly in and out lend themselves to the creation of harmonic pads and elicit a slower and more relaxed style of play. Since they fade in, users do not perceive small errors in the rhythmic timing. On the other hand, timbres with percussive attacks tend to elicit faster and more rhythmic styles of play. With these timbres, small timing errors, including the unavoidable timing errors caused by tracking latency, are easily perceptible and detract from musicality and user perception of control, since it is difficult to play notes in-time with the background drum patterns. Certain timbres contain internal rhythmic elements that are automatically tempo-synced to the global *pdMusic* tempo. While these rhythmic timbres ensure playing in-time with the background drum pattern, they are repetitive and can become monotonous over time. The same is true for the background drum patterns. Once set by the user, a drum pattern will repeat until manually stopped. For this reason, non-users in the area may find the music generated by *pdMusic* uninteresting. However, the drum patterns do serve

their main purpose, which is facilitating musical improvisation. It accomplishes this by acting as a complex metronome, helping users play in-time; also, it overlays a metric grid allowing users to make better-timed musical gestures and more accurate repetitions. For the purpose of spurring musical improvisation, the repetitiveness of the drum patterns is actually beneficial. The beats are complex, giving improvisers substantial musical material to play off of, but the pattern repetitiveness allows users to quickly learn each beat and compose innovations around it. A non-repetitive drum pattern could lead lead to instances of confusion where users craft improvisations for a pattern that has unexpectedly changed. Electronic dance music is one of the more repetitive genres of music. Usually, its rhythms change only slowly and subtly over the course of a song. Therefore, the repetitive aspects of *pdMusic* can be thought of as a factor that helps place the activity afforded by *pdMusic* into the genre of making electronic dance music.

5.1.3   Pilot Study

My third and final priority as designer of this system was that it serve as a platform for study on the topic of volitional usage and engagement. Therefore, I incorporated two subsystems into *pdMusic* facilitating this sort of study in the context of long-term installation. The first subsystem is an anonymous login system that affords the tracking of usage. In this system, using *pdMusic* requires a login and password. Acquiring this account information can be accomplished within minutes from the *pdMusic* interface—potential users submit a desired username and valid email address, after which a unique password will be automatically sent to the email address. This login system is essentially anonymous; users need not provide their real names at any point. Email addresses, however, are stored so that *pdMusic* can automatically resupply any forgotten passwords if requested. Once users have their account information, they can log in and use *pdMusic* at any time. *pdMusic* registers the date and time, the duration,

and the unique ID of the account holder, upon each usage session. Therefore, repeat usage is tracked, and can be used in quantitative calculations regarding frequency-of-use.

The second subsystem automatically gives short survey questionnaires to users while they play *pdMusic*. Each time a composition is played or recorded, a survey question pops up that is selected from a larger list of surgery questions according to the following scheme: the first seven survey questions asked are selected at random from a preliminary list given in appendix B.1. These questions are designed to elicit variables that should be controlled for, such as the extent to which each user has engaged with video games or musical activities in the past. All subsequent survey questions are selected at random from a larger list given in appendix B.2. These questions are designed to ascertain levels of player engagement, and obtain overall player impressions on the game's design. Answering questions is not mandatory. Questions from the non-preliminary list may be asked multiple times; this affords comparison between user responses over time.

In a pilot study of *pdMusic*, I installed it into the student lounge at Stauffer B, a building at Arizona State University, during the period between April 10, 2012 and May 9, 2012. In Stauffer B, undergraduate classes are held on topics related to Digital Culture, such as video-game design, media editing, and electronic music. Therefore, this location supplies a large number of potential users, with interests that are in line with *pdMusic*. The system's existence was promoted through class announcements, posters and flyers, and a text-based screen-saver on the monitor of *pdMusic* giving instructions on how to begin.

Table 5.1 summarizes the results of that pilot study. This was my first time seeing the system in use, and it proved instrumental in refining my design framework. My hypothesis was that most students with classes in Stauffer B would try the system, and that some fraction of those who tried would use the system repeatedly, in between classes, during the month of installation. Instead, as seen in table 5.1, only eleven people tried the system, and only one was a repeat user. Analyzing the cause for this difference between expectation and reality led to

169

Total Number of Participants in Pilot Study	11
Number Introduced during a digital culture showcase on April 20	6
Repeat Users	1 (with 2 re-engagements)
Users Introduced to System by a Friend	1 (introduced by the repeat user)
Number of Music-Makers	2 of 3 respondents (includes the repeat user)
Number of Video Gamers	2 of 2 respondents
Number of Movement Artists	1 of 2 respondents
Number of Hobbyists	1 of 1 respondents
Age Range (if Known) of Participants	21–29 (this question was asked 3 times)
Total Number of Questions Asked of Users	19
Number of Questions Answered	14
Mean Words Used per Answered Question	6.07 words per answer

Table 5.1: Relevant data from long study.

many changes in my design framework. Most notably, the impetus to develop a sub-framework for shorter-term volitional usage (as discussed in section 3.1) can be traced to this pilot study. Previously, the design framework consisted of only the longer-term framework, but this pilot study illustrated that any design framework for longer-term volitional usage that does not also deal with shorter-term volitional usage is faulty.

Therefore, the design framework was updated. Analysis of the data in table 5.1 gave insight while developing some of the shorter-term heuristics. For example, more than half of the total people who tried the system did so on April 20, because there was a public showcase of student experiential media projects on that day in Stauffer B, and many attendees of that showcase tried *pdMusic*. This begged the question of: Why did showcase attendees exhibit more volitional usage than students? The answer was clear—the showcase attendees had already budgeted the cost, in time and effort, of participating. They were at a showcase for interactive media (indicating some interest in the topic) during which there was no better alternative to trying the various interactive media systems, including *pdMusic*. On the other hand, students

who were at Stauffer B on a day-to-day basis were subject to numerous outside pressures and distractions. During the time-frame of the pilot study, many students were preparing for finals. Those who had spare time may have preferred activities that were more personally convenient and engaging, such as listening to music or surfing the internet. Also, the installation of *pdMusic* in the student lounge—although this was the only available space in the building—may have introduced problems related to demonstrability. For example, students may have avoided the system out of a sense of politeness to other students, as the student lounge is frequently occupied by students who are studying or napping. Or, if other students were present but not sleeping or studying, potential users may have foregone usage out of performance anxiety. The showcase setting on April 20 would have reduced these social effects because the purpose of the event was the demonstration of interactive systems. In short, *pdMusic* was not particularly convenient or well-targeted to the students of Stauffer B, but was convenient and well-targeted to showcase-goers. This fact instigated the inclusion of convenience and targeting heuristics into the shorter-term framework.

Before this pilot study, convenience and usability were part of my design framework only insofar as they were encapsulated in the heuristic of immediacy. Therefore, performing this pilot study helped me construct design theory with regard to these heuristics that encompasses the entirety of a usage career. For example, the login and questionnaire subsystems were designed for ease of use, but still added to the convenience cost and time cost of volitional usage. The login subsystem was problematic for convenience, requiring first-time users to perform several extraneous steps (submit a login request, check email, and login) before *pdMusic* could be tried, and also requiring users to memorize a password and username for any repeat usages (password and username information could be recovered if forgotten, but this required redoing the three original steps). This may have been a significant factor in why so few users tried the system. The questionnaire subsystem caused a usability problem by distracting users and removing them from the state of flow during play. It is possible, if students

had gotten past the first usages of the system, that they would have considered these issues of convenience and usability to have been minor in comparison with the engagement provided by the afforded activity, and longer-term volitional usage would then have ensued. This never took place, however, because few people passed the first usages. From this, I realized that convenience and usability are important in shorter-term volitional usage.

Soon after the pilot study was initiated, I performed a small usability test using three expert interactive systems designers as subjects. This resulted in a number of minor fixes to the system design which were easily implemented. However, two problems of greater severity were noted by all three testers that could not be fixed during the course of the pilot study. Each had to do with the use of the Kinect sensor. The first was that the Kinect sensor exhibited intermittent latency and tracking glitches, perceptible in the movement of the hand cursors. The second was that users were forced to switch between methods of controlling the system: arm movements control musical play via the Kinect, but between compositions, mouse and keyboard is needed to make selections and configure the system. All three experts felt that the system would be improved if all interface elements, including composition selection and system configuration, were controllable using arm movement, because switching modes of control has the effect of being annoying and removing users from the state of flow.

Based on this pilot study, *pdMusic* was redesigned and repurposed to create a second design iteration. This pilot study was also instrumental in guiding the selection of evaluation methodologies to be applied on this second iteration. The second iteration and the evaluations therein are described in the next section.

5.2 Second Iteration of *pdMusic*

In the first iteration of *pdMusic*, two of my foundational design approaches led to a system design impeding volitional usage as discussed in section 5.1.3. These two foundational design approaches were: a) sacrificing naturalness—and, by proxy, convenience and usability—for the purpose of data collection, and b) designing for longer-term volitional usage without considering shorter-term volitional usage.

In correcting the first design approach, the automatic survey questionnaire subsystem was disabled; it was at this point that the more holistic qualitative research paradigm, discussed at the beginning of this chapter, was adopted. The login and usage-tracking subsystem was also disabled, to be replaced with a more automatic anonymous login subsystem that works by detecting users in the Kinect field-of-view. Users are automatically logged out upon leaving the Kinect field-of-view, or can be manually logged out if they wish to cease usage but stay within the field-of-view. This updated login subsystem records the times of logins and logouts, but cannot distinguish between users. As the second iteration is designed for qualitative data collection, distinguishing between users is not a necessary feature of the login subsystem.

In correcting the second design approach, the application of the system was repurposed. Since many of the shorter-term framework heuristics are sensitive to the setting of usage, and *pdMusic* is good at facilitating the generation of electronic dance music, it was repurposed for installation at the sort of public events that might benefit from a DJ. This resolves many setting-related issues from the first iteration, such as the inappropriateness of the student lounge with regard to demonstrability and convenience. Due to the repetitive nature of its music, however, the first iteration of *pdMusic* seems a poor substitute for a DJ. Therefore, a human DJ—Andrea Silkie, who DJs as Sonique des Fleurs but is referred to as "the DJ" in this chapter—was incorporated into the system as a collaborator. She plays music that *pdMusic* users can improvise to (see appendix A), and all audio from *pdMusic* is routed through

her DJ mixer; this gives her volume control and allows her to apply audio effects. A drop-down menu for musical key was added to the *pdMusic* interface, allowing her to manually keep *pdMusic* in key with the background music. She also handles *pdMusic* timbre configuration, to ensure that sounds blend with her background music, and composition selection. Therefore, all mouse-and-keyboard interaction has been taken out of the hands of users; as discussed in section 5.1.3, the mouse-and-keyboard interaction decreased perceived usability in the first iteration. The drum pattern feature of *pdMusic* remains operable and under her control, but is frequently not used, otherwise it might clash with the background music. Similarly, the ability to record new compositions remains operable, but in the second iteration compositions are recorded only by myself, not users; composition recording requires skill with the system, so it is not an appropriate feature for one-time usage.

5.2.1 Modified Heuristic Assessment

Heuristic evaluation is performed by giving testers a set of heuristics, having them use a system a few times, and then asking for heuristic-related comments on the system. In Nielsen and Molich (1990), this method is used to identify potential usability problems and to suggest the significance of aforesaid problems. Greater significance is implied when multiple users independently comment on the same thing; if a usability problem is identified by only one person, this may indicate a personal preference unique to that person, but if a problem is identified by all testers, this suggests a severe and obvious problem. In the context of usability, three or four testers are usually sufficient to identify more than half of potential design problems; adding more testers produces diminishing returns with regard to identifying design problems, but remains helpful for determining the significance of previously-discovered design problems.

I held a demonstration of *pdMusic* for peers, during which I performed a modified heuristic evaluation. This modified evaluation used eight interactive media designers as testers. The purpose of this assessment was to identify usability issues and other potential imped-

iments to volitional usage and engagement. Instead of giving testers a set of heuristics to prompt observations—at this stage of my work, it seemed inappropriate to assume the validity of chapter 3 heuristics to this extent—I informally interviewed each tester after the usage session. The interview questions were loosely guided by aspects of the design framework. Interview subjects could answer questions with as much length or brevity as they desired, and were also allowed to launch into tangential ideas or thoughts. Due to the free-form nature of the interviews, there was no strict pattern, but certain questions recurred. Some questions that were frequently asked were:

- Was there anything about this experience that limited your ability to become engaged?

- Did you feel like you got better at this system as you used it? Do you feel there was room for improvement? If you used the system again, might you further improve?

- Did you feel the system offered enough variation or things to do to keep you from getting bored?

- Was it more fun to improvise music or to play compositions? Why?

- How did having this audience here watching impact your experience? Did it feel awkward at all?

- If this system were installed at your home, how often would you use it? Why?

- Do you have any suggestions for improving the design of this system?

I took short-hand notes of what was said. Data analysis was performed by dividing each notated interview into separate observations. Next, I grouped the observations into categories by topic, resulting in five topic categories: usability, incrementality, explorability, environment, and aesthetics. In cases where observations could have been coded into multiple topic categories I tried to choose the most relevant category. After compiling the topic categories across all interview subjects, I grouped within-category to find observations made by multiple interview subjects expressing the same things or overlapping things. Observations were paraphrased, combined, and given a significance value indicating the number of interview subjects making the same observation. These paraphrased observations, and the significance

values therein, are shown for each topic category in tables 5.2 to 5.6. Nested observations are used in cases where observations overlap but differ slightly. The following paragraphs discuss the topics and observations in more detail.

Observation	Sig.
Tracking needs improvement	7
...specifically, tracking lag	5
...and if tracking cannot be improved	
...rhythmically quantize notes	2
...spatially quantize hand cursors (i.e. snap-to)	2
...use timbres with slow attacks	1
...make targets larger	1
Visual indication of best distance from Kinect should be added	3
...without this, targets seem out-of-reach	2
Calibration time should be reduced	2
System text is hard to read	1

Table 5.2: Usability observations (and significance expressed as x testers of 8 remarking) from preliminary assessment.

Usability: Criticisms about frustrating aspects of the system were coded into the usability topic. This evaluation indicated that motion tracking was problematic. As seen in table 5.2, tracking problems were observed by seven testers, indicating that only one person found the tracking acceptable. Lag in the motion tracking made it difficult to control the spatial accuracy of hand cursors; the lag often resulted in the overshooting of targets, and five users observed that this frustrated their attempts at musicality. Assuming the lag was unavoidable, these users suggested alternative means of affording musicality without good hand control, such as: a) using rhythmic or spatial quantization to increase the musicality of the output since users are incapable of well-timed input, b) using timbres with slow attack envelopes because these do not inspire rhythmically-timed play, or c) making the targets larger so less spatial accuracy is required. In addition to the tracking problems, several other usability problems were mentioned.

176

Two users observed that calibration took too long. In the case of one of these users, however, calibration was delayed because a previous user had not left the Kinect field-of-view. A user observed that there was no visual indication of how far to stand from the Kinect. If such an indication existed, this would also have helped two other users, who attributed difficulty reaching targets to their short bodily heights, when in fact they could have reached more easily by moving closer to the Kinect. Lastly, one user found the system's text difficult to read.

Incrementality: Observations related to challenges inherent in the activity afforded by *pdMusic*, or to learning, were coded into the topic of incrementality. As seen in table 5.3, there was widespread agreement that this system affords learning and improvement both within single sessions and, in all likelihood, across multiple sessions. Of the five users who observed learning in single sessions, one noticed improvement at playing the system by feel, and deemed it possible, with practice, to play *pdMusic* without looking at the interface. Another user observed that more efficient playing styles developed over time. This user discovered arm sweeps and improved movement trajectories, because at first, the gestalt of the radial note-paths caused her to return to the center of the space between targets. Two users, however, observed that the basic concepts of the system could be understood after only a few minutes of usage—after this introductory process was finished, the pace of learning slowed because further learning required exploration of musical gestures and playing techniques. Lastly, one user framed the tracking issues seen in table 5.2 as a learning challenge. This user deemed it possible, with practice over multiple sessions, to anticipate the lag in the movement tracking in order to play in time with the music better.

Five early-stage challenges were observed. These challenges, unlike usability issues, seem both surmountable through learning and specific to the predilections of individual players. The benefits of reducing these challenges would, in all likelihood, be offset by drawbacks. One user observed that non-musicians might feel intimidated by the musicality required by this

Observation	Sig.
System affords learning and improvement	
...within single sessions	5
...probably, between sessions	2
Examples of learning include	
...improvement at playing by proprioception	1
...discovery of arm sweeping and other efficient playing techniques	1
Learning occurs primarily in first few minutes	2
Tracking issues in table 5.2 surmountable through learning	1
Some observed early-stage challenge barriers are	
...that non-musicians may feel intimidated	1
...that the time between composition notes is too small	1
...that wrong notes are possible to play during compositions	1
...that certain compositions cause arm crosses	1
...that targets not well-aligned for arm sweeps	1
pdMusic entry bar is higher than *Dance Dance Revolution*	1

Table 5.3: Incrementality observations (and significance expressed as x testers of 8 remarking) from preliminary assessment.

system; for this reason, this user felt that *Dance Dance Revolution* was better-designed for beginners. Another observed that the alignment of the targets is not well-matched to arm length, making wide arm sweeps more challenging. Three other users observed challenges specific to the playing of compositions. One felt that the duration between composition notes was too small, making it difficult for beginners to keep up. Another user suggested that notes outside of a composition should be muted so users cannot accidentally play wrong notes. Another noticed that certain compositions call for crossed-arm playing, which is more challenging.

Explorability: Feature requests asking for increased control methods, and observations related to feelings of boredom, were coded into the topic of explorability. As seen in table 5.4, five subjects felt that the system was not complex enough for users to avoid boredom for longer periods of time. Two users felt that the system configurations should change every few

minutes in order to avoid boredom; one of these users felt that the timbre setting, in particular, needed to change frequently for user interest. Another user, while discussing hypothetical future events, observed that he would personally try the system again if there were new timbres and new background music to explore in each new event.

Observation	Sig.
More explorability needed to prevent longer-term boredom	5
Users with musical training require greater operational freedom	6
...such as an ability to manipulate timbres	2
...potentially via the center-of-motion indicator	2
...potentially via foot or leg tracking	1
...such as an ability to transmit cues to DJ	1
...through an improved musical interface	
...color-coding the targets	1
...adding a visual history of recently-played notes	1
For variety, aspects of system should change	3
...every few minutes, with regard to configuration or song changes	2
...especially the timbre setting	1
...per event, new songs and sounds should be introduced	1
Tracking issues in table 5.2 reduce explorability, cause boredom	1

Table 5.4: Explorability observations (and significance expressed as x testers of 8 remarking) from preliminary assessment.

One user observed that more advanced musicians would desire a more complex interface, and this observation is supported by the fact that five other users, all with training in music, made feature requests that would result in greater musical control. Some of these users wanted to configure their own timbres, or wanted to have timbral control of certain features of the sound synthesis such as distortion or filter cutoffs. Of these, two wanted timbre control either via the center-of-motion indicator, or via foot or leg tracking. One user wanted the ability to give cues for the DJ to respond to. Another user had two suggestions improving the visual interface for musicians: the first was to color-code the targets to simplify the mental models and memorability of melodies, and the second was to implement some sort of visual history of

the recent targets that were played. For instance, playing a target could change its color slightly in a way that accrues, so that playing it repetitively causes the color to change more deeply. The target could slowly fade back to its original color if unplayed. This might help users better understand their musical gestures.

Lastly, one user observed that the problems with motion tracking detailed in table 5.2 have the effect of decreasing perceived control and operational freedom—i.e. explorability— thereby causing users to grow bored more quickly.

Environment: Observations related to the setting *pdMusic* is demonstrated in, including observations pertaining to the presence or absence of an audience, were coded into the topic of environment. As seen in table 5.5, there was unanimous agreement that the presence of an audience did not negatively impact usage of the system during this assessment. Three users did not observe any positive impact either—when questioned directly, two denied any feelings of awkwardness or performance anxiety and moved on with the interview, and the third implied that the audience was arbitrary by suggesting that the system would be equally engaging if used at home alone. Of the other five users who observed a positive impact due to the audience, two felt that the audience increased engagement, and three observed a motivation to perform well for the audience. Of these three, one user observed that this motivation to perform well reduced her desire to practice repetitively or explore aimlessly while improvising in the system. Only two observations were made expressing reservation about the potential impact of audiences on user engagement, and both were discussing hypothetical situations. One user felt that if the system were hypothetically modified to demand more musicality, non-musician users might feel awkward in the presence of audiences. Another user observed that if the system were hypothetically demonstrated at an event where those in attendance did not know each other, some users might avoid the system because of performance anxiety.

Observation	Sig.
Audience improves experience	5
...audience motivates performing well	3
...and reduces desire to aimlessly explore music	1
...audience increases engagement	2
Audience causes no feelings of awkwardness or performance anxiety	3
...presence of audience is arbitrary to experience	1
In hypothetical future events, audience might depress volitional usage	
...if event-goers are unknown to one other	1
...among non-musicians, if the system demanded more musicality	1
If *pdMusic* were hypothetically installed at an event attended by the user	
...the user would be likely to try it again	2
...assuming the setting was appropriate for *pdMusic*	1
Hypothetically, if *pdMusic* were installed at user's home	
...it would elicit lengthy practice sessions	2
...it would not be used without the social context	1

Table 5.5: Setting-related observations (and significance expressed as x testers of 8 remarking) from preliminary assessment.

Many observations were based on hypothetical environments defined by users. For example, in hypothetical future events, two users observed that they would personally use the system again, although one of these users qualified the statement by saying the setting would need to be appropriate for *pdMusic* usage. In other words, the event would need to contain party-like or club-like elements befitting an electronic music DJ system. Otherwise, if the environment was not well-matched, users would be unlikely to try it. This user recommended a music festival such as Coachella[6] for *pdMusic* installation. In a hypothetical installation directly into the homes of users, two users claimed that they would personally take part in lengthy practice sessions; one of these two users stated that this would give her a chance to simultaneously participate in three personally enjoyable activities, namely exercise, musical improvisation, and active listening to the background music. On the other hand, another user observed that he

---

[6]see http://www.coachella.com/

would be unlikely to use *pdMusic* if it were installed at his home. For him, the social context and convenience of the demonstration event provided the impetus to try the system, and this would be lost at home.

Aesthetics: After the previous topic groups were formed, all remaining observations were commentary either pertaining to specific system features, or speaking to the level of overall fun in the system. These were grouped into a catchall topic labelled 'aesthetics.' As seen in table 5.6, three users observed that the system was fun to use, one of whom called it both fun and exploratory. More users preferred playing existing compositions over improvisation than the reverse. One user observed that the feedback for playing notes and compositions correctly is satisfying. A user noticed that some of the timbres were monophonic, and this user observed a preference for polyphonic timbres as these afford the playing of chords. Lastly, one person mentioned that the system was well-designed for busy locations and hang-outs, mainly because the music is unobtrusive enough that conversations can take place around it.

Observation	Sig.
Playing compositions is more engaging than improvising	4
Improvising is more engaging than playing compositions	2
Feedback when playing compositions correctly is satisfying	1
Polyphonic timbres preferred over monophonic	1
Using system is fun	3
...and exploratory	1
Music is unobtrusive and appropriate for busy settings	1

Table 5.6: Aesthetic observations (and significance expressed as x testers of 8 remarking) from preliminary assessment.

Outcomes of Preliminary Assessment

This preliminary assessment was successful in identifying areas of improvement in the design of *peMusic*, and resulted in various changes being made to the system. Observations that implied

potential design improvements were each analyzed for benefits and drawbacks, making sure to factor in the time required to implement the improvement. The following paragraphs discuss design decisions that were made based upon the observations from this preliminary assessment.

The most significantly observed usability problem (as seen in table 5.2) was noise and lag in the motion tracking. Observations suggest that these tracking problems have repercussions in the heuristic of explorability as well as usability, and this is a useful insight with regard to my design framework: it stands to reason that usability problems and responsiveness issues have the effect of impeding user agency and operational freedom. The *pdMusic* tracking code was refactored at great length to reduce the lag and noise, but neither could be eradicated entirely as these problems are, in part, byproducts of using the Kinect for sensing (all other forms of optical motion capture contain lag and noise as well, to varying amounts). Therefore, to make musicality easier in the face of noise and lag, the targets were made larger to reduce the need for pinpoint movement accuracy, and the timing of all generated note-ons and note-offs were quantized to a sixteenth-note metric grid. I acted on several observations that were of low significance in cases where potential benefits outweighed potential drawbacks and implementation was not too time-intensive. For example, the targets were colorized to better afford musical understanding and mental models, and also aligned on an elliptical ring in order to better afford arm sweeping. System text was redesigned for improved readability. After these changes, the modified visual interface is shown in fig. 5.7. This figure can be compared to fig. 5.3.

Of course, many of the observations in this preliminary assessment were offered in the spirit of commentary and do not imply any call for design changes. Some other observations were acted upon by changing protocols for system demonstration, rather than changing the system design. For example, rather than implementing an indicator in the feedback telling each user where to stand, it was decided to place a line of tape down at an appropriate standing distance, knowing that the DJ and I would be able to provide additional direction for extraor-

183
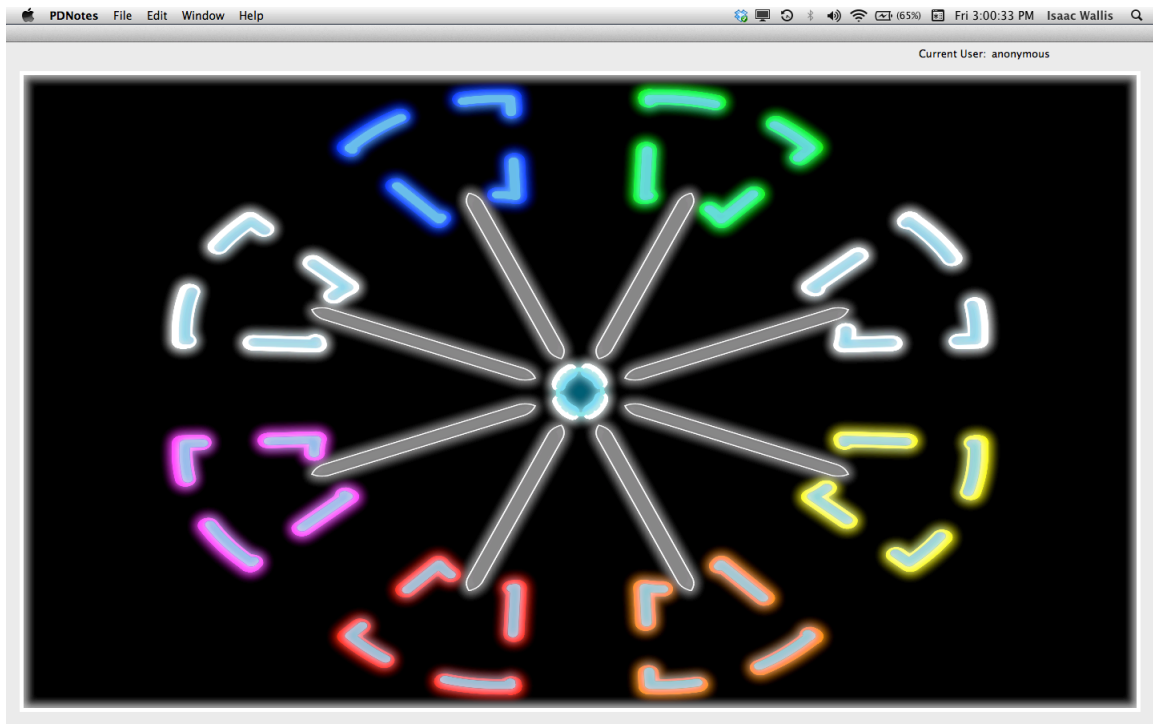
Figure 5.7: *pdMusic* interface redesign based on heuristic assessment.

dinarily tall or short users. Based on observations, it was decided to emphasize timbres with slow attacks during demonstrations if tracking problems were taking place. Also, it was decided that the DJ would change system configurations, timbres, or background tracks approximately once a minute in order to increase variety.

Many observations, on the other hand, were not acted upon because the potential drawbacks or time required for implementation seemed to outweigh any potential benefits at this time. For example, observations calling for simplifying challenges at the cost of musical freedom (i.e. disabling unneeded targets during compositions to preclude the playing of wrong notes, redesigning compositions to avoid crossed-arm play) were disregarded. All musical instruments contain these sorts of technique-related challenges, and overcoming them can be an engaging activity. However, other observations calling for increased musical complexity were also disregarded, as the anticipated public events for *pdMusic* installation would be attended by many non-musicians, and I did not want *pdMusic* to intimidate these potential users. Lastly,

184

some observations were not acted upon at this time because this iteration of *pdMusic* is more focused on shorter-term volitional usage than longer-term. These may be acted upon in future iterations. For example, one called for the ability to send cues to the DJ using the interface. Implementing this design change may allow me to better understand the heuristic of cooperation. However, as cooperation is more of a longer-term heuristic, this was not implemented.

5.2.2   Deployment

pdMusic was installed and demonstrated at five art events in the local Phoenix metropolitan area. During these events, the DJ ran *pdMusic* with occasional assistance from me, and I conducted users, talked with them, and collected data. Both of us had very busy roles. In my case, I was required to: a) perform various tasks in the effort of keeping the system running, b) explain the system to potential users, c) hold conversations with previous users and notate their observations, and d) observe and notate my impressions of user behavior, audience behavior, and the behavior of other event-goers. Needless to say, managing all these tasks simultaneously is not possible, so tradeoffs had to be made for the purpose of controlling my cognitive load. Since note-taking drew my attention from potentially significant user behavior, I eventually settled into a style of data collection where observation or conversation were interrupted as infrequently as possible; this meant that often I would wait until lull periods to make detailed notes. This data collection style has the downside of relying on my memory, but was the only way to minimize oversights while honoring the guiding research tenet of naturalness in user interactions. Otherwise, even if I scribbled observations in short-hand, user behaviors would often be overlooked and conversations with users would become stilted as I would need to take frequent pauses for note-taking.

The evaluation being performed at these demonstrations is, like all experimental study performed at Arizona State University, subject to ethical review by the institutional review board (IRB). As seen in appendix C, the IRB exempted my evaluations from review as long

as users under the age of eighteen are excluded from data collection (i.e. they are allowed to use the system, but it is not allowable to collect data about them), and the anonymity of all participants is maintained (i.e. it is not allowable to collect data that identifies participants or that could be used to identify participants in the future). These restrictions disallow the use of video recordings for data collection; however, after the first two events, I requested permission from the IRB to use audio recording to ease my data collection process with voice notes and recorded interviews. Permission was granted provided I erase any utterances of user names from the audio recording upon my first listen.

Often, qualitative researchers have preliminary knowledge on the sorts of behaviors they will be observing, allowing the creation of data collection and data analysis protocols beforehand. I did not have such preliminary knowledge, unfortunately. Therefore, during the first three public events I used a flexible data collection protocol to be performed for each user, using the data collection sheet shown in fig. 5.8. This sheet's fields are described as follows:

- Subject No: The first user of the night is annotated here with index 1, the second with index 2, and so forth.

- Date: The date of the public event is annotated here.

- Time Started: The time of day when usage is started is annotated here.

- Time Stopped: The time of day when usage is ended is annotated here.

- Time Observing (before, after): The approximate amount of time a user spends observing the system, both before and after usage, is annotated here.

- Would Use Again: If, in conversation with a user, the question is asked: "Would you use this system again, if it were installed at some other public event and you were there?", the answer is annotated here.

Subject No._____Date:_____Time Started:_____Time Stopped:_____
Time Observing:_____before_____after
Would Use Again?_____
Focused Attention:_____with:_____
Affect:_____

Notes:_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Figure 5.8: Subject data collection sheet for first three events.

- Focused Attention (with): The extent to which users seemed absorbed in the activity, and any descriptive indications used to make this observation, are annotated here. According to O'Brien and Toms (2008), focused attention is a characteristic of engagement.

- Affect: Any observations with regard to the user's demeanor or affective state is annotated here. Change in affect is a characteristic of engagement according to O'Brien and Toms (2008).

- Notes: Any observations not appropriate for other fields are annotated here.

As can be deduced by the disproportionately large size of the 'Notes' field, this data collection sheet was designed with miscellaneous observations in mind. After three events, the collected data was submitted to a process of open coding and axial coding (to be discussed in section 5.3). This resulted in data categorized into sixteen axial codes; these codes were used to redesign the data collection and coding protocols, including a more detailed data collection sheet as shown in fig. 5.9. In this new sheet, every field correlates with one of the sixteen axial codes. These codes are described in appendix D.2.

In sections 5.1.3 and 5.2.1, the installation environment had great impact on volitional usage. One significant contribution of this dissertation could be to increase our understanding of how system setting impacts volitional usage. Therefore, just as observations about the users are coded and treated as data, observations about the installation events are coded and treated as data. The code categories are described in appendix D.1, and the following paragraphs describe each event in detail, including selected remarks made by users at each event.

First Friday at Space 55:   There is a district in Phoenix where several independent art galleries and other arts-related businesses are located. On the first Friday of every month, all of the galleries hold an event where entrance is free. This event has become one of the most well-attended monthly events in the area; the district gets flooded with thousands of attendees going

Subject No._____Date:_____Gender:_____Age:_____
Attended Because:_____Accompanied By:_____

Time Started:_____Duration Played:_____Number of Sessions:_____
Time Observing: before_____after_____Reason Stopped:_____

Sociability_____Anxiety:_____

Focused Attention:_____with:_____
Vocalizations:_____
Learning:_____Rhythmic Movements:_____

Knowledge/Interests:_____Locus:_____
Would Use Again?_____
Time Used in Comparison with Other Works?_____

Comments:_____
_____
_____

Criticisms:_____
_____
_____

Compliments:_____
_____
_____
_____

Notes:_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Figure 5.9: Subject data collection sheet for events subsequent to the first three.

through all the galleries. Many of these attendees consider themselves part of various artistic counter-cultures; often these attendees sport distinctive fashion statements or hairstyles. On the street during this event, bands play, break-dancers perform, street preachers proselytize, and a wide variety of costumes are worn. The age range of attendees skews toward younger adults.

I viewed this event as a source of potential users who may be especially receptive to a novel musical system like *pdMusic*. Therefore, on First Friday November 2, 2012, I installed *pdMusic* in Space55, which is an independent theater venue located about three blocks from the main First Friday event[7]. Typically this theater hosts modern plays with edgy or adult themes. My system was operational from 7:30PM to 10:00 PM.

For the bulk of people coming to the area for First Friday, the location of Space55 is somewhat inconvenient because it is several blocks from the main First Friday district. I hoped that users, while walking to the First Friday district, would pass Space55 and would then come in and try *pdMusic*. Parking is often hard to find during First Friday, resulting in heightened foot traffic in the area. Since a small art gallery next door was also holding a free event, I also had hopes that some people attending that event would stop in. To encourage this, a sign was placed on the sidewalk outside of Space55's entrance describing the system and welcoming users. Also, I sent a promoter—Kristi Rogers—to the main First Friday district to pass out event flyers and describe *pdMusic* to potential participants. The design of the promotional materials is shown in appendix E. In addition, the event was publicized using email lists and Facebook.

In this event, there were seven users, all male, four of whom had no previous exposure to the system and were unknown to anyone associated with *pdMusic*. Three of these users were attracted to the space by flyers or interaction with the person who was handing out flyers. The fourth one stopped in while passing by Space55. Of the other three users, two had no previous exposure to the system but were known to the DJ or I. The third was known to me and had

---

[7]see http://www.space55.org

tried the system on one previous occasion. The data from these last three users is still valid, but it is important to acknowledge that friendships or other interpersonal relationships may have contributed to their volitional usage. Because of the low number of users, it was possible to construct a timeline of usage for this event, which is given in appendix F.

This was the first public showing of the system, and as is often the case, situations came up during the event that were not considered when programming, such as multiple audience members in the background while users played with the system. These situations exposed bugs in the system leading to periodic crashes, some of which were severe enough to require complete restarts of the computer. In subsequent events, these problems were largely ironed out. In consequence, this event was uniquely subject to usability issues and convenience issues arising from software bugs.

Family Day at the Arizona Museum of Youth: Twice yearly, the Arizona Museum for Youth[8] (AMY) holds a Family Day where the museum is free to enter. AMY is a small museum, mainly focused on exhibits designed to capture the interest of children and adolescents. It was decided for ethical reasons that data would not be collected on people under age 18, but many adults also attend AMY in order to participate in family-based activities. According to estimates by staff members at AMY, approximately 400 people attend the museum on typical Family Days.

I installed *pdMusic* into AMY on a family day which was held on November 3, 2012. The system was operational from 10:00 AM until 4:00 PM excepting a one-hour break for lunch. During this time, there were 21 users (4 male, 17 female). None of these users had previous experience with the system or any association with people involved in *pdMusic* development. I viewed this event as providing an interesting contrast with the previous installation at Space55: this installation was more accessibly-located for event-goers since it was located right in the

_____

[8]see http://www.arizonamuseumforyouth.com

museum, but many event-goers attended for reasons related to family togetherness or providing learning opportunities to their children—so *pdMusic* was not particularly well-targeted to this group. The result was a relatively large number of users, but the typical user exhibited less engagement and participated for a shorter period of time. Many were distracted by the requirements of dealing with their children. On the other hand, this event did offer many opportunities to observe the role of purpose and social interaction in activity participation.

Selected User Profile—Subject 3: This user was a staffer at AMY who worked there throughout the day. Like most other users from this event, she did not use the system for a great length of time. The reason subject 3 was interesting was because of the number of times she re-engaged. She tried the system a total of five times over the course of the day, with no single usage exceeding a duration of five minutes. One might conjecture that she chose not to use the system long because she was supposed to be working. There was no indication that the system was especially well-targeted for this user, and there was no opportunity to interview her as she always hurried off to work after each session. For this user, the extreme convenience of the system's setting led to her trying the system whenever she had a few spare moments as she moved from place to place within the museum.

Second Friday at Mesa Arts Center: On the second Friday of every month, the city of Mesa hosts an event that is inspired by Phoenix's First Friday in an effort to increase public engagement with the arts. Several area galleries and museums hold events, and vendors and musicians take to the streets of downtown Mesa. This event is not as well-attended as Phoenix's First Friday, and many attendees are local residents. However, this event still sees attendance numbering in the high hundreds to low thousands. Mesa Arts Center[9] (MCA) is located near the district where this event is held, and frequently holds theatrical or musical performances

---

[9] see http://www.mesaartscenter.com

during this event. Mesa Arts Center has a campus the size of one city block containing multiple stages and locations for creative performances.

On the second friday of November 9, 2012, *pdMusic* was installed in an outdoor location where event foot traffic was likely to be high. This outdoor installation afforded the use of louder speakers and an optical projector, so the visual interface was larger than on the previous events. The installation was operational from 6:00 PM until 10:00 PM. During this time, there were thirteen users (nine male, four female). Four users were were previous users of the system, and were acquaintances of the DJ or I, and three others were acquaintances but had no previous exposure to the system. There was light rain toward the end of the event, but it was not severe enough to force an end to the installation.

In terms of convenience and targeting, this installation fell between the Space55 and AMY events. The location was not central to Second Friday but was close enough that many Second Friday event-goers would pass by. Although Second Friday event-goers as a whole may not be as interested in experimental art and music as the users seen in the Space55 installation, they are also less likely to be distracted by dealing with their kids than the users seen at AMY. As a result, the number of users in this event fell between the numbers in Space55 and AMY, and the persistence and engagement exhibited by users also fell between that exhibited at Space55 and AMY. The city of Mesa has ordnances against promoting events by passing out flyers on the street, so we promoted this event by putting flyers in local coffee shops.

Selected User Profile—Subject 2: Subject 2 happened upon the system installation along with two other friends. At the time his group came along, subject 1 was still enthusiastically using the system. Subject 1 was social and demonstrative in her usage of the system, laughing and joking with passers-by as she learned to use the system. In conversation with subject 2 and his friends before his usage session, one person asked me what the system was for. When I replied that it was being used in a study on user engagement, another person joked: "She *[subject 1]* sure likes it!" Subject 1 stopped using the system soon thereafter—I believe she

saw that other potential users were waiting—and subject 2, of the three friends I had been conversing with, quickly agreed to try the system.

Subject 2 displayed an intense degree of focused attention over a relatively long time span (38 minutes). This duration exceeds any other usage session across all events. His friends left after watching him play for approximately ten minutes, but subject 2 did not notice. During the time spent playing, this user's style of play changed significantly. For example, toward the beginning of his session, when he played existing compositions he would start notes at the correct time, but not sustain them the correct length. By the end, without any prompting from us, he had learned to play through all notes.

After thirty-eight minutes, he looked at his cell phone, then asked: "How long does this game go?" When I explained that it would be running continuously for several hours, he said "Oh. Well, I gotta go." I asked if I could walk with him in order to ask a few questions, and he agreed. The transcript of the short interview goes as follows:

Me: If this system were installed at a public event and you just happened to be there, would you be likely to use it again?

Subject: (*emphatically, as if surprised by the question*) Yeah!

Me: Great! Also, did you have any thoughts on the system you'd like to share with me?

Subject: Not really, It's just something really new, to me anyway.

Me: Okay, well thanks for answering my questions.

Subject: No problem. I'm just mad my two friends decided to skip out on me while I was playing. That's why I had to stop.

Selected User Profile—Subjects 4 and 9: This interaction highlighted an interesting interplay between the intrinsic motive of purpose and the heuristics of promotion and targeting; namely, the role of word-of-mouth. Subject 4 happened upon the system in the course of regular Second Friday activities while walking with two of his friends. He used the system for about sixteen minutes while his friends observed. During this time, I had a conversation with his friends where I learned that all three were workers at a local hacker-space which is walking

distance from the MCA. This group of people was knowledgeable about programming and interactive media. About an hour after this group left, subject 9, another worker at the hacker-space, showed up to try the system. Subject 4 and friends had returned to the hacker-space and given a glowing review of the system, inspiring subject 9 to come try it. This was an encouraging example of volitional usage, indicating that the original group found *pdMusic* interesting and well-targeted. It also suggests a self-selection process implicit in word-of-mouth promotion: subject 9 was well-targeted by the system because friends (who were also well-targeted by the system) originally made him aware of it.

Slipstream Art Exhibition on First Friday at the Icehouse   On Jackson street in down-town Phoenix, about 1.5 miles from the First Friday district, there is a venue called The Ice-house[10]. The Icehouse is a historical building, an actual icehouse where ice was manufactured in the 1920's. This venue has been used for arts exhibition since 1990. On the First Friday of December 7th , 2012 (from 6:00 to 10:00 PM), an art event called Slipstream was organized by Mary Neubauer, in part as an exhibition for her students in the sculpture department of ASU. Several of the art pieces were interesting and novel interactive installations. This represented a good opportunity to exhibit *pdMusic* in an environment where the novelty of the system could be isolated. That is, since all of the art on exhibit was novel, and much of it was interactive, the usage of *pdMusic* could be studied in an environment where (unlike in the previous events) it was not the most novel system in the location.

Despite this, *pdMusic* had a crowd of somewhere between four and ten onlookers throughout most of the night, making it one of the more well-attended pieces of the event. The physical location of *pdMusic* within the Icehouse was convenient for event-goers, because we were set up in one of the main entrances to the building, right next to a refreshment table. The event was already promoted, so we did not promote it further. However, when I noticed

---

[10]see www.theicehouseaz.com

that many people were traveling through the space without using the system, I changed *pdMusic* settings in order to show the silhouettes of people in the space. This was meant to grab the attention of passing potential users, and it was effective in its goal. In fact, many people became engaged with dancing and posing while watching the silhouettes, even while other users were controlling the music. It seemed to draw many more potential users, and more observers, than there might otherwise have been. This highlights an interesting facet of the heuristic of promotion; namely, that carefully designing an interface to catch the eye of potential users is a form of promotion.

System calibration was a little finicky at this event, and for the last hour of the event we were forced to compete sonically with a drum circle. This proved impossible, so we ended up giving the DJ a break and turning off the background music, letting users improvise along with the drum circle instead of her electronic music. As it turned out, this still afforded user engagement with the system, and we continued accommodating users until the end of the event. Over the course of the event, we had 26 users. Many of these used the system quickly then moved on within a few minutes. However, some users desired longer periods of interaction. Of these, the most interesting is subject 10.

Selected User Profile—Subject 10: This user provided a great deal of usable data. He observed the system for a long period of time before he used it for the first time. He used it for only two minutes, because the Kinect tracking lost him as a user:

Me: I'm sorry, the Kinect lost you!
Subject: *[jokingly]* I was just about to get 'on fire'!

There was no system crash, and the system was still operational, but this subject refrained from using the system for more than thirty minutes, during which time he left the area. Then he tried again. This time, he persisted for eight minutes. This subject was highly interactive with the audience, joking with them and showing them different uses of the silhouette

196

feature. Examples include doing a breast-stroke, doing Michael Jackson dance moves, and pulling audience members in front of the Kinect to shadow-box with him. He was also affable with me and other people who were there helping the system run. After doing a pose, he jokingly asked my note-taker: "Did you get that?" During the time he played with the system, the crowd of observers began to grow in size. Some of these observers may have stopped to watch him play, and others may have stopped to see what was drawing such a crowd.

Subject 10 used the system a total of four times, for a total duration of 31 minutes, longer than all but one user in the five events. He made it clear that the reason he took breaks from the system was so that other potential users would have a chance to play. For example, after his second session, he said "I would play more, but I want to give others a chance to play." Later, on another session, he exited the system while apologizing to another prospective user with "Sorry, I'm a screen-hog." However, at one point a girl came up to the system and expressed an interest in playing it. She asked him several times what to do, then danced in front of the Kinect to watch her silhouette. Subject 10 was happy to answer her questions, and told her how to use the system, and even danced next to her. Interestingly, though, he did not step aside to let her play. At this point he seemed so engaged with the system that the thought may not have occurred to him. After a few minutes, the girl left to see the rest of the art projects.

After his third session with the system, I was able to strike up a conversation with him:

Me: I think I may know the answer to this question already, but if this system were installed in a future setting and you were there, would you try it again?

Subject: Yes. Yeah, I think what it is—I don't really get into the dancing that much—it's more just trying to psychologically do it. It's like a brain-teaser, trying to get the movement. And then you have have to get into the music too, to do it right... but yeah, I've barely left here all night!

Subject 10 highlights an interesting fact about *pdMusic*; namely that extraverted personality types may find the system more enjoyable than other people. This would stand to reason, given the fact that *pdMusic* is installed publicly where audiences are free to observe.

First Friday at monOrchid: The events at Space55 and Icehouse each took place on a First Friday, but *pdMusic* was not centrally-located to the First Friday activities in either event. The typical First Friday audience seemed especially well-targeted by *pdMusic*, and many users exhibited engagement and volitional usage in those previous events. Therefore, it seemed important to observe usage on a First Friday with an installation location that was central to the event. On the first friday of February 1, 2013, I installed *pdMusic* into the art gallery monOrchid[11]. MonOrchid is one of several art galleries on the main First Friday thoroughfare, therefore its foot traffic during First Friday is very high. On the evening of the installation, it also featured paintings by a local artist. Some of these paintings had sexualized adult themes, so signs were posted to warn parents lest their children be inadvertently exposed to the painting content. In this event, I collected data on 24 users. It is important to note, however, that this event was an outlier: its audience size was at least double that of any other event of *pdMusic* installation. At the heaviest part of the night, it became impossible to record data quickly enough to keep up with user turnover. Therefore, it is likely that the actual number of adults who tried the system is closer to 35 or 40.

---

[11]see http://monorchid.com/

5.3   Analysis

My work on the framework in chapter 3 was already well underway by the time I began creating *pdMusic*, allowing me to use the framework as a theoretical lens during data analysis. As discussed in section 2.2.3, the approach to theoretical lenses taken in O'Brien and Toms (2008) was based on coding data into groups according to components of the theoretical lens—in their case the components were the 'threads of experience' from McCarthy and Wright (2004)—resulting in a final analysis containing themes centered around those components. In my situation, however, I was using my own framework as a theoretical lens, so taking the O'Brien and Toms approach could lead to systematic confirmation bias. In other words, by using my own theories to explain user behaviors, I risk missing alternative theoretical explanations that are superior in terms of fit, relevance, workability or modifiability.

To minimize this risk, I combined the O'Brien and Toms approach with grounded theory methodology. This does not eradicate the possibility of confirmation bias in my data analysis—this is not possible within the context of any qualitative research—but does keep the bias from being built directly into the research protocol. Per Creswell, pg. 13, grounded theory is "a strategy of inquiry in which the researcher derives a general, abstract theory of a process, action, or interaction... using multiple stages of data collection and the refinement and interrelationship of categories of information." In other words, grounded theory is a systematic qualitative research procedure that is used, not for theory testing as in quantitative research, but for theory generation. Data analysis in grounded theory consists of multiple coding passes. The first pass is 'open coding,' where a brief summary is made for each line of data, and each summary is made without consideration of the other lines of data. Typically, this results in a large quantity of open codes, tantamount to the original number of lines of data. The second pass is 'axial coding,' where themes and categories are identified among the open codes, thus reducing the number of variables. The next pass is 'selective coding,' where the most relevant

and predictive variables are identified; these selective codes are the beginnings of the grounded theory, and should be analyzed for fit, modifiability, and workability using all new incoming data. In this data analysis, I performed open coding and axial coding, then analyzed the axial codes for their relevance to framework heuristics.

In addition to the data analysis given in this section, two supplemental interview transcripts are given, to be found in appendices G.1 and G.2. The first transcript is a long interview with the DJ performed on Dec 17, 2012 over the course of 45 minutes. As my most involved collaborator and longest-running system manager, she had valuable insights on the system. The second is a shorter interview with an expert user taken immediately after her first usage session on Feb 1, 2013 during a lull period at the monOrchid event. As an unbiased HCI expert with no background in music or gaming, her insights were similarly valuable.

Event	Users	Time
Space55	7	11:25
AMY	22	4:46
MCA	13	9:50
IceHouse	26	4:42
monOrchid	25	4:21
Total	**93**	**5:51**

Table 5.7: Number of users and average usage time per user (in minutes and seconds) in each *pdMusic* installation event.

Data was recorded on a total of 93 users. As seen in table 5.7 and fig. 5.10, volitional usage varied widely between events. As discussed in section 5.2.2, the events were each marked by unique circumstances impacting volitional usage. The most important selective codes are discussed in paragraphs below. The data addresses some heuristics more thoroughly than others. This is to be expected, as the framework is large and complex; it is unlikely that any single system or study could support or inform all framework heuristics.
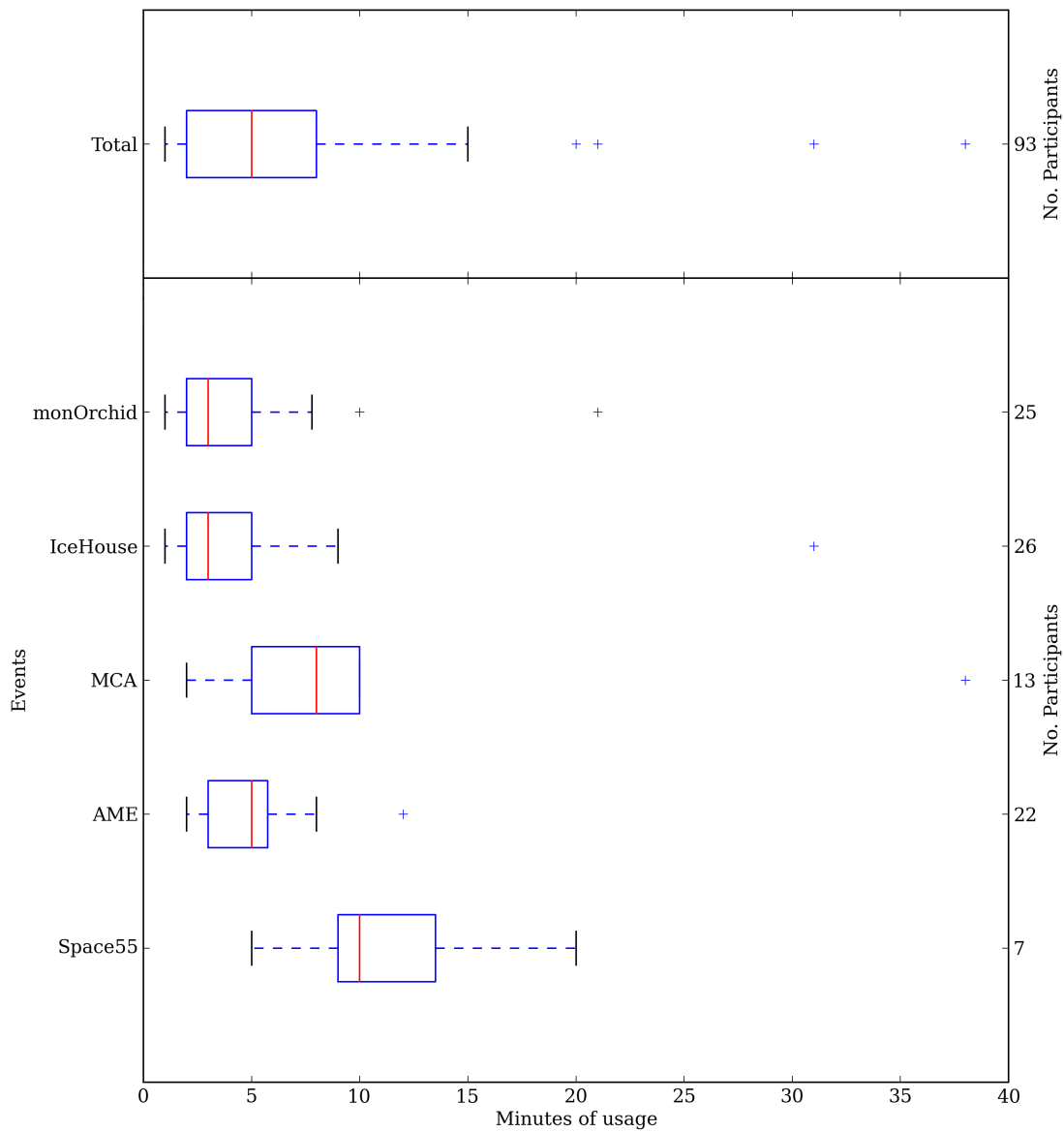
Figure 5.10: Usage per event. Red line is the median usage time, the box plot spans the quartiles, the whiskers span 1.5 IQR, and the crosshairs represent outliers.

Convenience: The convenience of *pdMusic*, as an interactive system installed at public events, seems to be reducible to three factors:

- The number of potential users attending the public event, because visiting *pdMusic* is more convenient for nearby event-goers;

- The cost, in terms of time or effort, of visiting the *pdMusic* installation for event-goers;

- The incidence of distractions, outside opportunities, or outside pressures associated with trying the system for those at the *pdMusic* installation.

Table 5.8 provides my observations on that data. Exact tabulations of attendance were impossible to obtain in these chaotic installation environments, so the figures presented are approximate and relative judgements. Similarly, costs of attendance and usage for individual attendees are unknowable, but observable time and effort costs that were event-wide are listed.

Event	Wider event	Wider event size	Attendance costs	*pdMusic* attendance	Usage costs & distractions
Space55	1st Friday	> 5000	walking 3 blocks	≈ 15	crashes
AMY	Free Family Day	≈ 100	watching kids	≈ 80	watching kids, other exhibits.
MCA	2nd Friday	> 100	walking 1 block, rain	≈ 25	rain
IceHouse	Slipstream	≈ 150	none for wider event attendees	≈ 150	system availability, other exhibits, drum circle.
monOrchid	1st Friday	> 5000	navigating crowds	> 300	system availability, other exhibits.

Table 5.8: Information relevant to the convenience of each event.

Costs of visiting the *pdMusic* installation should depress the number of wider event-goers that visit the installation, and as seen in table 5.8, this logical relationship was upheld in these installation events. The costs often took the form of walking distance. Space55 is several blocks away from the First Friday center of activity, and most First Friday event-goers walk from place to place, so First Friday event-goers had to walk that distance if they wished to visit the *pdMusic* installation—this could explain why such a small percentage of the First Friday crowd visited the *pdMusic* installation. MCA is not central to the activity of Mesa's Second Friday, but is within sight of the active area, so a slightly larger percentage of event-goers visited the *pdMusic* installation. *pdMusic* was installed centrally to the events at AMY and IceHouse, and almost all event-goers also visited the *pdMusic* installation in these events. *pdMusic* was also central to the First Friday activity when installed at monOrchid, but in this case a smaller percentage of event-goers visited the installation. This discrepancy can be explained by costs related to navigating crowds—it can be difficult to get in and out of galleries such as monOrchid on First Friday due to the amount of people in attendance.

As can be seen by comparing tables 5.7 and 5.8, there exists a positive correlation between the size of the audience attending *pdMusic*—the potential user base—and the number of actual recorded system users. However, as one might expect, costs of using *pdMusic*—including outside distractions, opportunities, and pressures—depress the number of *pdMusic* audience members who choose to try using the system. These costs took many forms, but one of the more common forms of cost were delays and outside pressures arising from the fact that only one person at a time can use *pdMusic*, so turns must be taken. This cost could help explain why a relatively small percentage of monOrchid's potential users chose to try the system—there was insufficient time for all of them to try the system. The same might be true of the IceHouse event. On the other hand, Space55 had a very small audience, but a relatively large percentage of that audience tried using the system. At that event, there was little need to wait for a turn.

203

Promotion: The Space55 event exposed a triadic relationship between system location accessibility, targeting, and promotion. Due to the distance of the installation from the activity-center of First Friday, flyers were distributed in the main First Friday area as a means of promotion. This resulted in a very small base of potential users, but a very well-targeted one. As explained in the discussion on convenience, location inaccessibility may have caused the base of potential users to be very small. However, the audience that did attend was extremely well-targeted: only those who were highly interested in the *pdMusic* concept were willing to pay the effort cost and walk to the installation. This explains, in part, why the users at the Space55 event persisted in usage longer than users at other events. During that event, the five people who were given flyers averaged 12:24 minutes of usage each; the remaining two, who attended because they were friends of mine, averaged 9:00 minutes of usage. In conversations, these five users revealed knowledge and interests in subject areas related to electronic music, computer programming, game design, or other areas related to *pdMusic*.

In most cases, the events of *pdMusic* installation were promoted by the event organizers (as in the case of the Slipstream Art exhibit at the Icehouse and the Free Family Day at the Arizona Museum for Youth), or were well-known regular events (as in the case of First Friday at Monorchid or Second Friday at Mesa Arts Center). The DJ and I promoted each event through our own social media accounts, and this resulted in some of our friends and acquaintances attending installations. Although the Space55 installation took place on a First Friday event, it was also promoted separately using flyers (viewable in appendix E). The 21 users who attended due to our system-specific promotion had a mean usage time of 8:53 minutes, while the 72 who attended for the wider event had a mean usage time of 4:57 ($p < .006$). Since there is a logical relationship between promotion and targeting, it is also worth noting that, of those who attended due to promotion specific to *pdMusic*, there was a higher incidence of previous relevant interests that were notated. Only nine of the 72 users who attended for the wider event claimed relevant interests, while 18 of 21 users who attended because of promotion specific

to *pdMusic* claimed relevant interests (the remaining three accompanied others who claimed relevant interests). This data is biased due to the composition of our social networks, but it seems likely that promoting a system makes it more likely that interested users will use that system, and those interested users will then be likelier to persist in usage sessions for longer durations. Refer to fig. 5.11 for a promotion-specific box-plot.

The heuristic of promotion also encompasses the way in which a system advertises itself by catching the eye of people in the area. At the IceHouse event, it was observed that more users tried the system once the system began displaying the silhouettes of all passers-by. The movement seemed to catch the eye of potential users, who would often spend some time, before system usage, observing their silhouettes moving in the system. In this study, efforts were made to notate when people spent time observing the system before usage. The 23 users who observed the system for at least two minutes beforehand had an average usage time of 9:08 minutes, while those who used the system with less than two minutes of preliminary observation had an average usage time of 4:21 minutes ($p < .001$). One possibility is that many users who found the system interesting enough to watch may have also been better targeted to the system, resulting in more persistent usage. Other possibilities exist, however, such as that users who spent time watching before-hand may have understood the system better when they used it, resulting in a more enjoyable experience.

Targeting: Many users made remarks stating or implying previous interests in activities that are related to *pdMusic* such as dance, media art, programming, or music. Some users explicitly stated these interests (i.e. "I'm a dancer, so I really love this system") and others made remarks implying these interests (i.e. "What programming language did you write this in?"). As mentioned above in the discussion on promotion, users who overcame significant time or effort costs in order to use the system tended to have previous interests that were aligned with the *pdMusic* activity in some way. The 27 users who had interests in related activities had an average
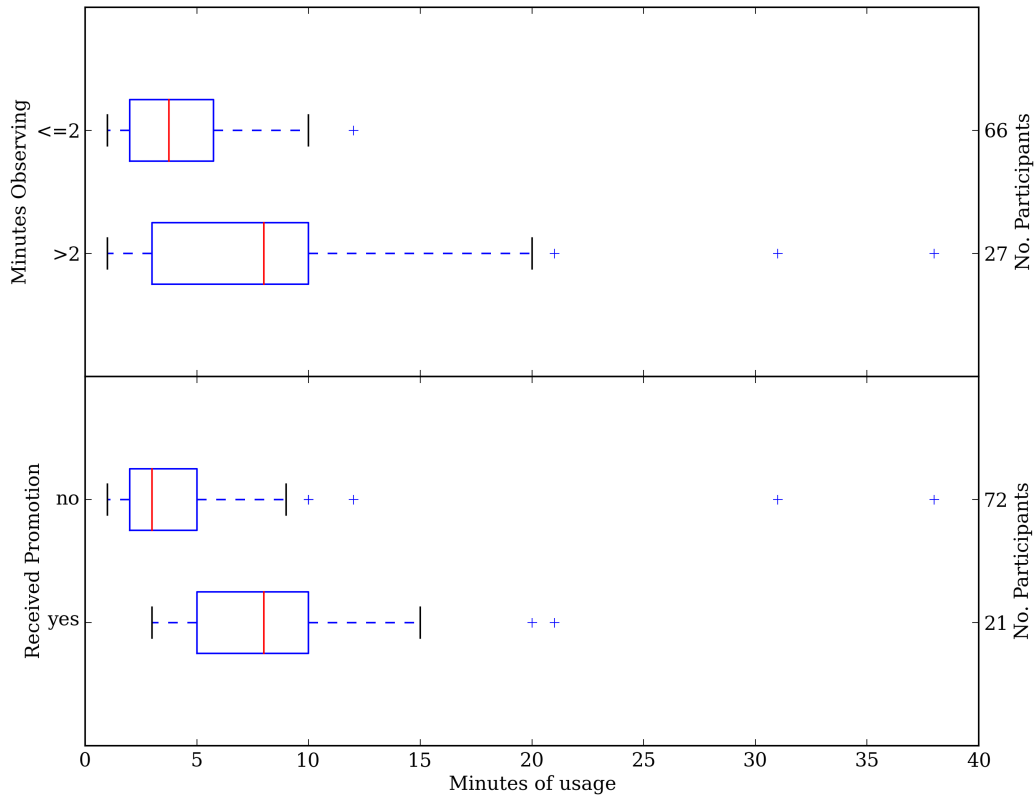
Figure 5.11: Data analysis relevant to the heuristic of promotion.

usage time of 8:08 minutes, while the remaining 66, without known related interests, had an average usage time of 4:55 minutes ($p < .014$).

Since the *pdMusic* installations took place in the context of wider events, another targeting-related factor is the match between the spirit of *pdMusic* and the ethos of the wider event. For example, in the AMY event, there is a clear mismatch between the ethos of *pdMusic* and the wider setting. Whereas *pdMusic* is a video-game-like interactive music system suited for dance clubs, parties, or experimental art spaces; the wider event held at AMY was focused entirely on education and providing activities for children. On the other hand, the Slipstream event that was held at Icehouse was very well-matched to the spirit of *pdMusic*, and many of the other

experimental art pieces that were exhibited also had interactive elements. Perhaps this is why, on a per-hour basis, the Icehouse event had more users (at roughly 6.5 users per hour) than the AMY event at (at roughly 4.4 users per hour) despite the fact that the installation was located centrally to the event in both cases.

Certain observations were made on user demographic traits—namely age and gender—which could, theoretically, be relevant to targeting. Age did not impact average usage time in any statistically significant way, and although there were 69 young users (i.e. users who appeared to be between the ages of eighteen and forty) verses only 24 users who appeared to be older than forty, this may have been due to the age composition of the wider events. Gender, on the other hand, did have an effect on usage time. There were 54 male users with an average usage time of 6:55 minutes each, while there were 39 female users with an average usage time of 4:22 minutes each ($p < .04$). These differences in gender may be attributable to factors built into the installation environments, but if so, they must have been common to all events, because the lop-sidedness of usage between genders did not significantly improve when looked at between events. One possible cause of this discrepancy is that *pdMusic* could be more appealing to men than women.

Usability: The idea that usability is an important factor in the size and dedication of a user base is not controversial; it is the basis for all studies on the topic of usability. Therefore, supporting this idea was not a focus in this evaluation, and there was not much data that was relevant to to the heuristic of usability. However, there were two observed phenomena with relevance to the heuristic of usability. The first is this: *pdMusic* had more bugs and crashes in its first installation at Space55. As outlined in appendix F, these crashes were so severe that they reduced the overall number of users, because the crashes happened as the users wanted to begin their sessions. The time required to reboot the computer proved too inconvenient for

207
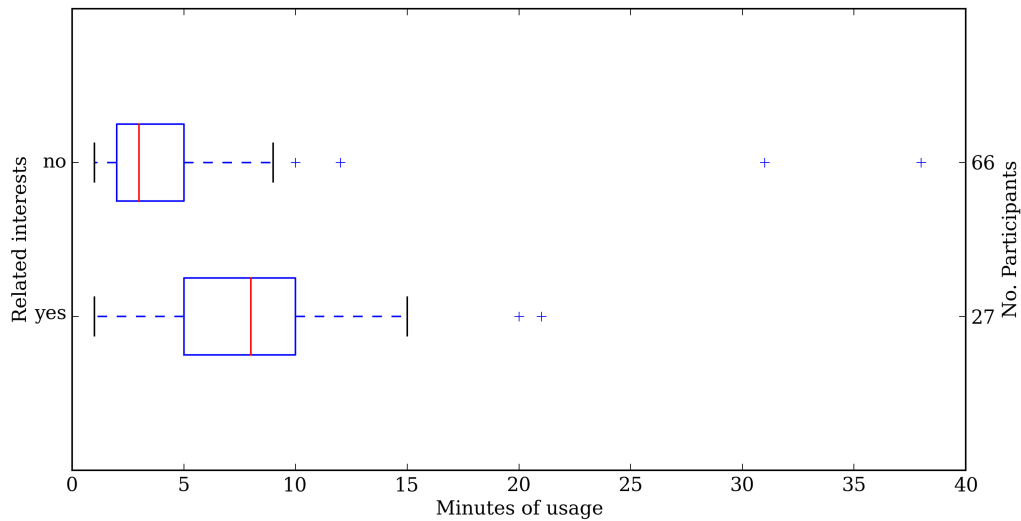
Figure 5.12: Data analysis relevant to the heuristic of targeting.

these users. Later, other people cut short their usage sessions after becoming frustrated with the crashing problem.

The crashing bug was eliminated after the Space55 event, but other bugs that were more difficult to solve persisted into later events. In particular, system tracking would freeze momentarily whenever the DJ made changes to the interface settings, and many users perceived this as a glitch. The DJ attempted to make all changes as quickly as possible; nevertheless, one pattern the DJ and I observed was that some users—especially those who were trying the system out of minor curiosity but were not highly engaged—would end their usage sessions upon encountering one of these glitches. Even though the freezes were only a few seconds long, it took very little to cause these users to quit usage.

Cognitive Richness: The heuristic of cognitive richness refers to the extent which an interface captures user attention, therefore it encompasses many topics such as aesthetics, fun, and challenge. There is no way to quantitatively measure how much attentional bandwidth an individual is using, but observations were made regarding apparent user focus levels. The four

levels of focus observed in this data are given in the list below. It is important to note that these levels of focus are not indications of where users were looking. All but the most distracted users kept their eyes on the system because the system is hard to use without looking.

- Tuned out: These users became engrossed to the point of ignoring their surroundings. They ignored questions and speech, missed social signals, and in some cases, failed to notice when their friends left them behind to go do something else. Of the 93 users, there were five who exhibited these behaviors, with an average usage time of 19:48 minutes.

- Focused but interactive: These users were focused on the system, but also interacted with others. They asked questions or made comments about the system, or they noticeably performed for their friends or the audience. There were 27 users who exhibited these behaviors. Their average usage session was 8:31 minutes long.

- Noticeably distracted: These users' focus was elsewhere, and the source of their distraction was apparent; for example, they may have been holding an unrelated conversation while using the system, or attending to a young child. There were only four users who were distracted to this extent. Their average usage time was 3:30 minutes.

- All others: Many users exhibited none of the above behaviors during their sessions. They may have been completely engrossed or mostly uninterested; their behavior gave no indication. There were 58 users like this. Their average usage time was 4:00 minutes.

Undirected user vocalizations made during usage were also notated, and these vocalizations are of interest because they tended to be either unconscious or rhetorical. Therefore, they were notated separately from commentary and remarks that are clearly directed at a listener or listeners. For example, if someone loudly exclaimed the phrase "This is awesome!", that would have been notated as a compliment, not a vocalization; but if someone said the word "awesome" under their breath while using the system, that would have been notated
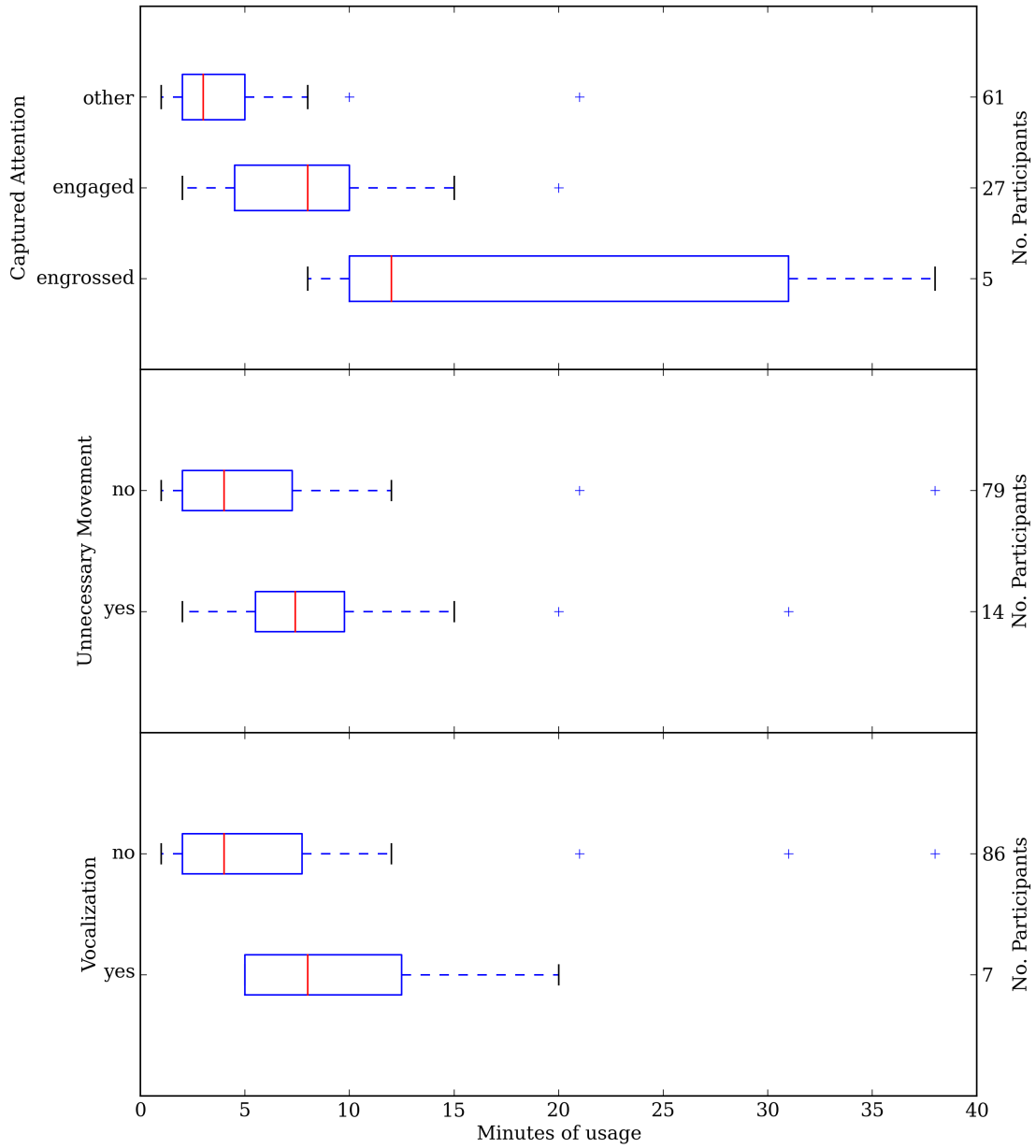
Figure 5.13: Data analysis relevant to the heuristic of cognitive richness.

as an undirected vocalization. Observed vocalizations were often exclamatory in nature, and included spontaneous laughter, screams, or grunts. Most vocalizations seemed unconscious, stemming from user engrossment, although it is possible that some users screamed or grunted in showmanship while being performative with the system. Seven users were observed making vocalizations, and they averaged 9:42 minutes of usage time; the remaining 86 users averaged 5:32 minutes of usage time ($p < .07$).

Similarly, any stylistic user movement that was clearly unnecessary for *pdMusic* usage was notated as observation. The vast majority of stylistic movement that was observed was rhythmic in nature. Some user movement was based on misunderstandings of interface affordances, as in the case of subject 2 from the Space55 event, who initially thought that *pdMusic* would track and respond to rhythmic movement. Some stylistic movement was performative or humorous in nature. Often, however, the stylistic movement seemed completely unconscious, emerging from user effort and engagement. Fourteen users had movement styles I found noteworthy enough for recorded observation, and they averaged 9:33 minutes of usage time; the remaining 79 averaged 5:11 minutes ($p < .01$).

Novelty: The heuristic of novelty is not well-addressed by the data, and it may be a challenging prospect to design studies that look at the impact of novelty. In the events described in section 5.2.2, a few users made unprompted remarks on the newness of the system, but any questions asked to tease out user perceptions of novelty invariably invited comparisons between *pdMusic* and other activities and systems with similarities to *pdMusic*. Certain user comments could have been interpreted as oblique references to system novelty (e.g. "What a great idea!"), but such interpretations are not definitive enough to base claims on. At this point, the impact of novelty is an open question.

Immediacy: The heuristic of immediacy, as defined in chapter 3, consists of time, effort, or financial costs leading up to participation. All *pdMusic* installations and wider events were

free of charge to attendants. Time and effort costs are covered elsewhere, in discussions on usability and convenience.
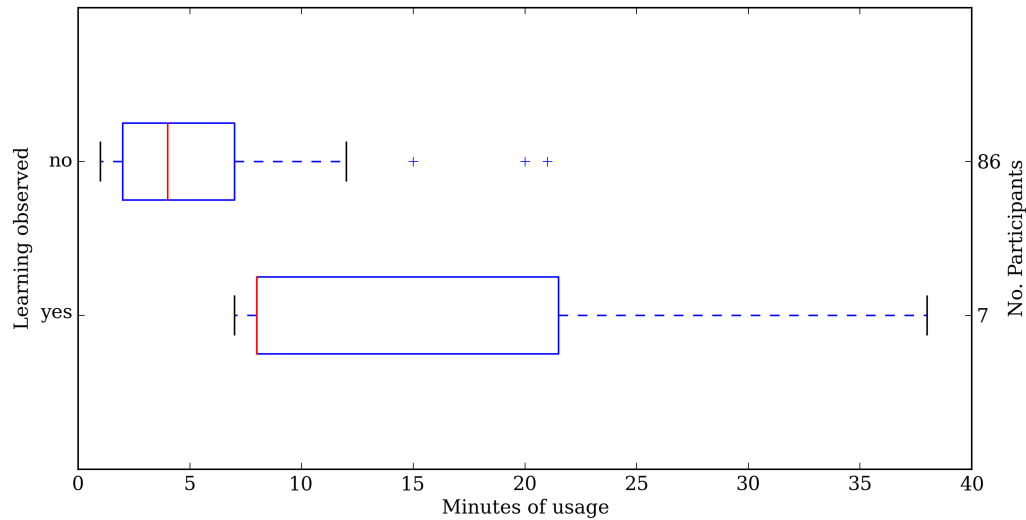
Figure 5.14: Data analysis relevant to the heuristic of incrementality.

Incrementality: In the context of this study based on short-term event installation, the impact of learning curve gradualness could not be ascertained. However, observations regarding learning during system usage were recorded, and the seven users with learning observations averaged 16:00 minutes of usage, while the 86 remaining averaged only 5:01 minutes of usage ($p < .001$). However, more observations were recorded about users who were under observation for longer periods of time, and this is a potential source of bias. To illustrate, the 46 users who participated in less then five minutes of usage had observations recorded in an average of 7.4 of the coded categories, whereas the 47 users who participated in five or more minutes of usage had observations recorded in an average of 11.4 categories. It was impossible to notate many recorded observations on users who quickly entered and exited the system, then left without conversation. This resulted in a Spearman correlation of 0.6 between the number of minutes spent in usage and the number of coded categories with observations

($p < .01$). Observations or user remarks on the topic of learning seem especially vulnerable to this bias. Since learning entails evolving behavior, it requires significant exposure and exploration, so observations of learning are an area where it is important not to conflate correlation with causation.

Ownership: As the heuristic of ownership is a longer-term factor, it was difficult to address in the context of this study based on short-term event installation. However, ownership is being exhibited when users take an interest in improving a system, and all user remarks, including feature requests and criticisms, were notated as observations. One potential source of bias in this data is that, since talking with me before or after usage implies some small expenditure of effort, all users who spoke with me were more engaged than users who did not speak with me. The 41 users who held conversations with me averaged 8:10 minutes of usage, where the 52 remaining averaged only 3:45 minutes of usage ($p < .001$). To eliminate this source of bias, the following comparisons based on user compliments, feature requests, and criticisms, are made only between users who engaged in conversation with me.

Many users complimented the system, but many of the complimentary statements were vague. For example, the most frequently received compliment was that *pdMusic* was 'cool.' As it happens, there was no significant difference in usage time between the 24 users who made compliments and the 20 users who conversed with me but did not compliment *pdMusic*. In contrast to compliments, criticisms were meaningful to volitional usage—but in a counterintuitive way. The seven users who made criticisms averaged 13:20 minutes of usage, whereas the 36 users who did not make criticisms averaged 7:01 minutes of usage ($p < .025$). One would think that users who perceive and criticize issues with a system might exhibit less persistence within usage sessions, but the opposite was the case. It is possible that users could only find these problems through exploring and thinking about the system at length during usage, leading to this result. On the other hand, another explanation might be that those who
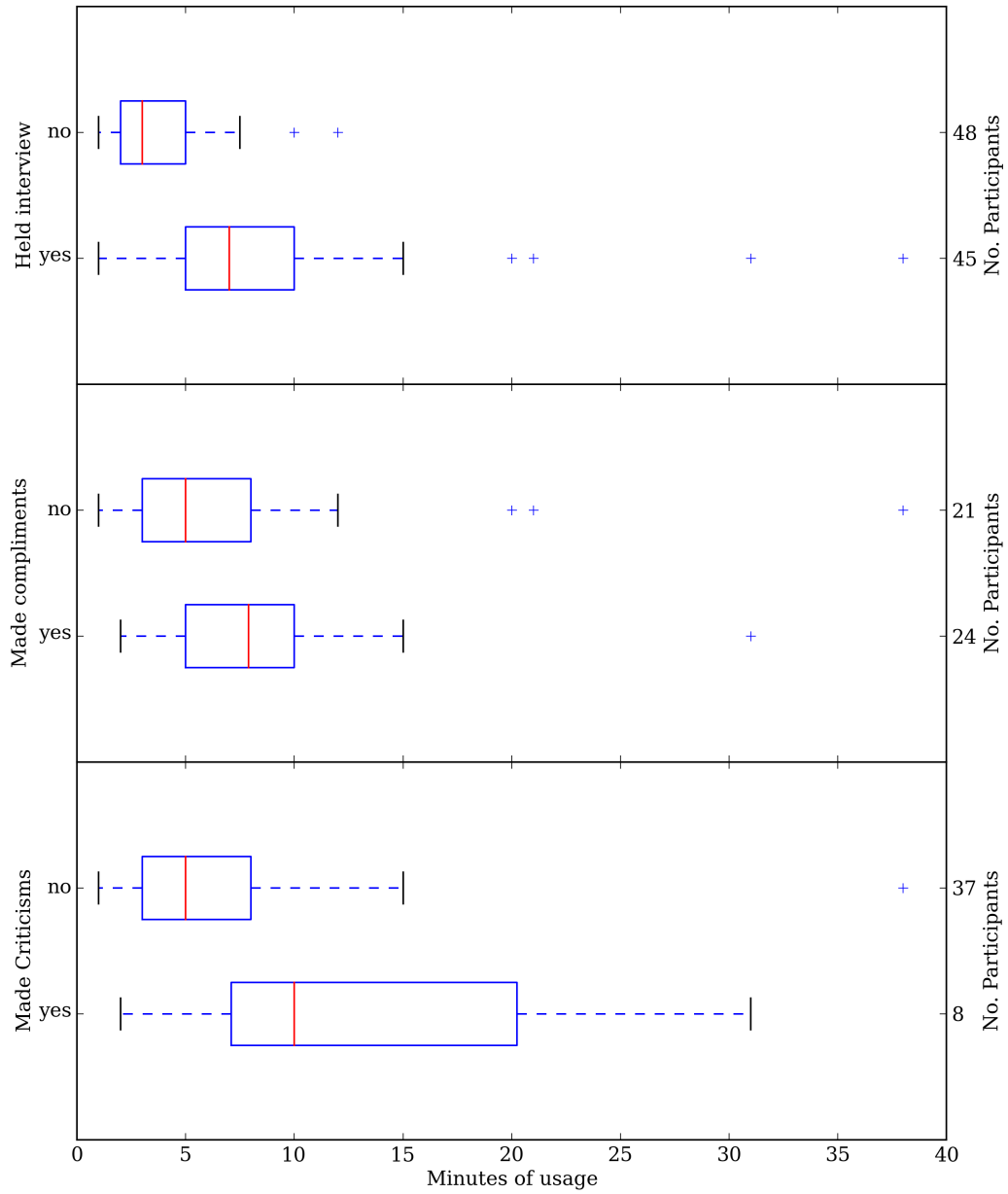
Figure 5.15: Data analysis relevant to the heuristic of ownership.

were being critical of the system were exhibiting a higher degree of system ownership by taking an interest in improving future iterations. Many users who made compliments on the system may have been merely acting out of politeness, with little interest in or desire to improve the system. Politeness is not a likely source of criticisms. The extent to which ownership drives the correlation between criticisms and longer usage sessions is not known, and could be difficult to extricate from the heuristic of targeting, since people that have a pre-existing affinity for some aspect of a system are more likely to identify dissimilarities between the system and the activity at the center of the pre-existing affinity.

Explorability: Because the heuristic of explorability is a long-term factor, it could not be addressed in this study based on short-term event installation. Therefore, the impact of explorability currently remains an open question.

Demonstrability: Any indications of user performance anxiety or enjoyment in performance were notated as observations. Indications of performance anxiety took the form of user remarks expressing feelings of awkwardness or initial reluctance. The eight users who expressed such feelings averaged 4:15 minutes per user, and the 83 remaining averaged 6:00 minutes per user, but this difference was not statistically significant ($p < 0.41$). Indications of enjoyment in performance took the form of notably high levels of audience engagement or sociability. There were seventeen users that exhibited these behaviors, averaging 10:07 minutes per user, while the remaining 76 averaged 4:53 minutes per user ($p < .001$). Of these seventeen sociable users, five treated *pdMusic* as a platform for performance, explicitly engaging audience members and trying to capture attention through humorous or exaggerated usage behavior. These five users were among our most engaged-seeming users, and they averaged 13:48 minutes of usage while the remaining twelve sociable users who did not behave in a performative way averaged 8:35 minutes of usage, although this difference was not statistically significant ($p < 0.2$). Since sociability and performative behaviors are related to extroversion, care must

be taken in making claims based on this data. Currently, two questions remain: First, would other systems offering similar performance opportunities draw a similar level of volitional usage from extroverted people? Second, if both extroverted and introverted users were made equally comfortable in their social surroundings, would the difference in usage persistence be reduced?
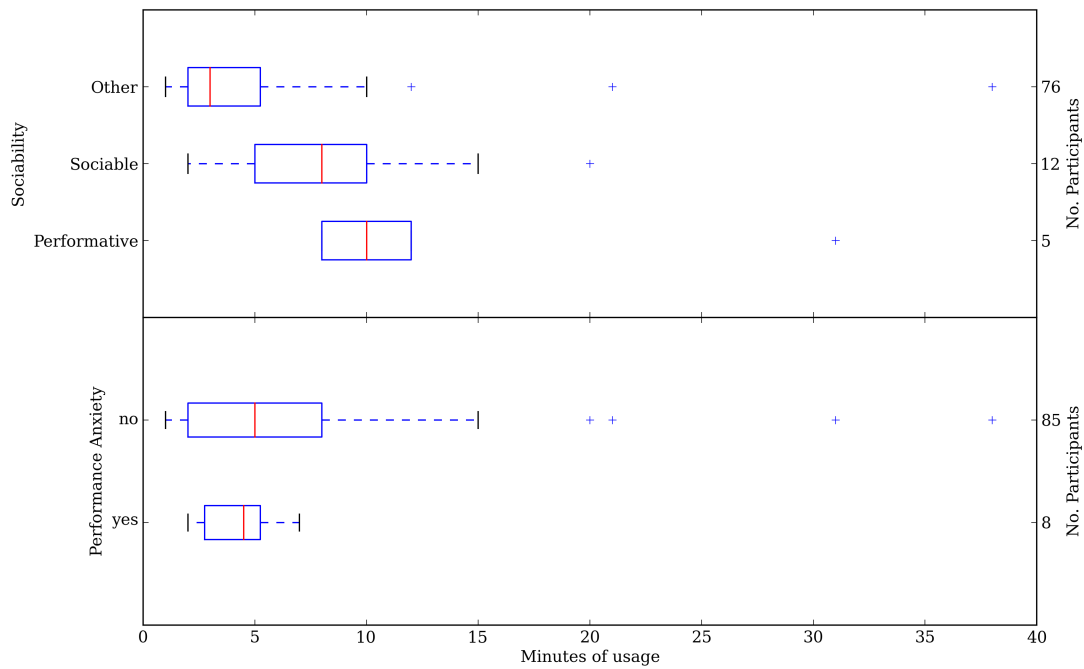
Figure 5.16: Data analysis relevant to the heuristic of demonstrability.

Cooperation: Comparing the evaluation from the second iteration to the evaluation from the first iteration, it seems likely that social considerations were a factor in the discrepancy between usage rates. In the first evaluation, where system usage could have been viewed as mildly impolite by students using the lounge, usage rates were far lower than in the second evaluation where experiencing art and trying interactive systems was part of the rationale for the public events. In section 3.1 I modeled usage session length as a function of user engagement

verses mounting outside pressures. Certain events had more built-in outside pressures than others, and often these outside pressures were social in nature. For instance, the monOrchid event had more people waiting to use the system, while the AMY event had many users who needed to attend to children. As seen in table 5.7 and fig. 5.10, events with built-in social pressures and distractions tended to have shorter usage sessions.

Between individuals, the reasons people had for attending the public events had implications on their persistence in usage, as did whether they came accompanied by friends or family. In some cases, outside pressures resulted in users taking multiple shorter sessions rather than single longer sessions. There were eight users who took multiple sessions. In two of these cases the secondary sessions seemed to be a result of users finding secondary usage highly convenient, meaning they had nothing else to do that they found more engaging. In the other six cases, outside pressures caused the users to end their initial usage sessions despite being interested in continuing usage, so subsequent sessions were undertaken. In five of these six cases, the outside pressures took the form of other audience members waiting to try the system. In the sixth case, the outside pressure took the form of employment—the user was at work and could only take short breaks to use *pdMusic*.

In this evaluation, it was observed when users attended the system in the company of others. Forty-two users came with friends or family, and these users persisted in system usage for an average of 7:27 minutes, while the other 51 users persisted an average time of 4:31 minutes ($p < .014$). Of the 42 users who came with a group, 21 were the sole group member trying the system, leaving 21 other cases where multiple group-members tried *pdMusic*. Interestingly, those who were the sole users from their group persisted longer—averaging 9:01 minutes—than those with peers also trying the system, who averaged 5:53 minutes of usage, although this was not a statistically significant result ($p < .2$). It seems likely that users with peers also trying the system shortened their usage sessions in order to be considerate and take turns. If so, this implies a politeness-based mounting outside pressure to end usage. This
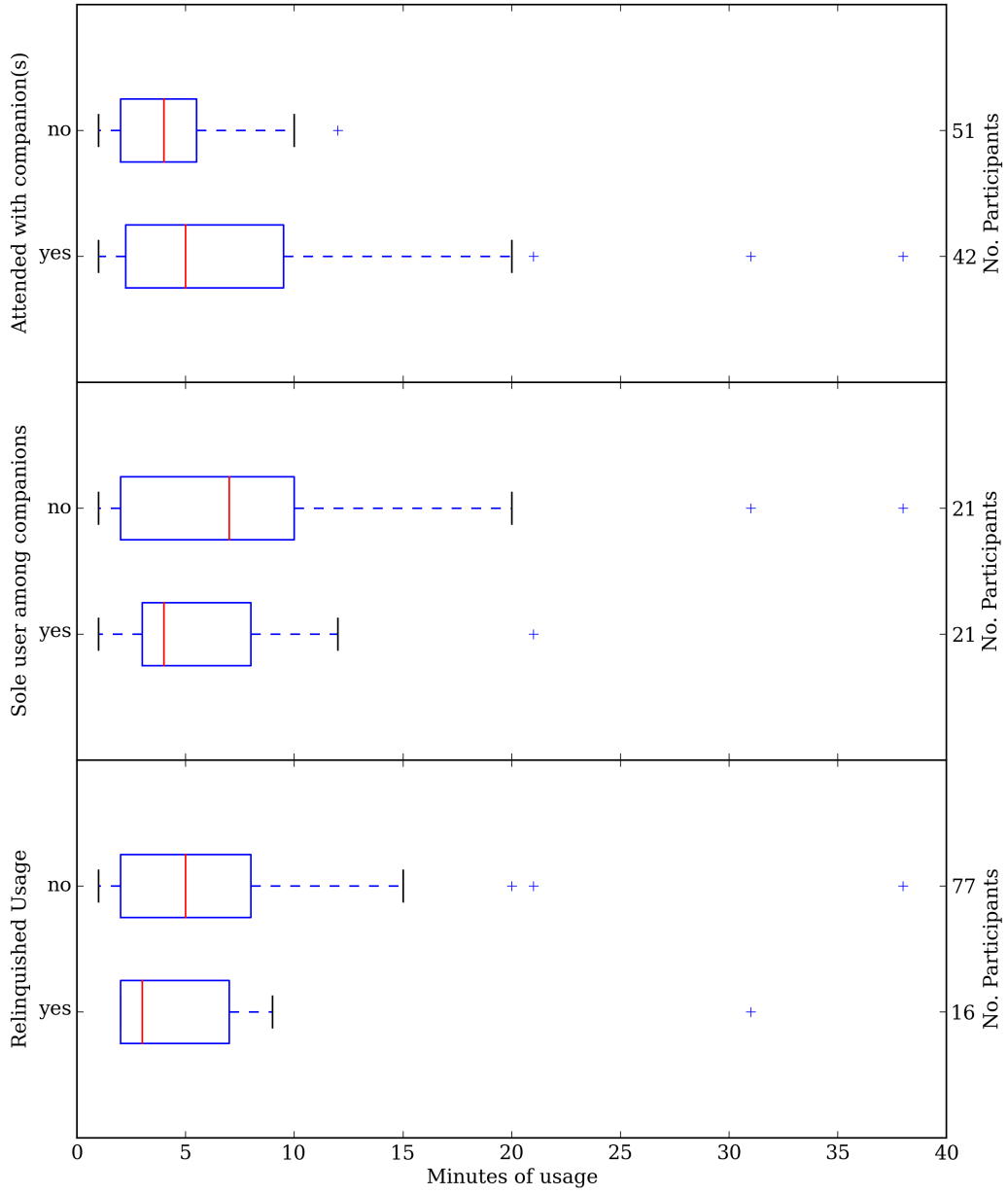
Figure 5.17: Data analysis relevant to the heuristic of cooperation.

implication is further supported by the fact that, as seen in table 5.7, there exists an inverse correlation between event audience size and mean usage duration. If users are aware that others wait their turn, they are more likely to cut their sessions short. To further explore this, an observation was recorded whenever it was apparent that a user relinquished usage in order to give someone else a turn. The fifteen users who clearly relinquished usage in this way averaged 4:03 minutes of usage. The remaining 77 users averaged 5:52 minutes of usage, although this is not a statistically significant difference $(p < .21)$. This statistic excludes one outlier, subject 9 of the IceHouse event, who took four usage sessions and relinquished usage on two of those; but did not relinquish usage on two other occasions when audience members hinted they were interested in taking a turn.

Chapter 6

FUTURE WORK

The work described in this dissertation comprises the beginning stages of a science of volitional usage, and much work remains to be done. The results discussed in section 5.3 are encouraging. They do not conflict in any substantial way with my design framework. All usage behavior that was observed can be explained using the framework. Surprising or counterintuitive observations informed my thinking on certain framework heuristics, but did not call for large framework modifications. Still, since this was only the first study on the topic of volitional usage, these results beg further exploration. For example, some of the trends identified in section 5.3 require further inquiry, because these trends seem generalizable to other experiential media interfaces; the likely causes of these trends are apparent, framework-related, and exist in many other interfaces. These generalizable trends are stated as propositions below. These propositions seem logically sound and do not conflict with existing data, but need to be tested further under divergent conditions, using systems that are unlike *pdMusic*.

Proposition 1. *If all else is held equal, but people must travel further or pay other immediacy costs to gain access to a system, the size of the potential user base will decrease.*

Proposition 2. *If all else is held equal, a greater number of potential users will result in a greater number of actual users.*

Proposition 3. *Users willing to pay upfront immediacy costs are likelier to persist during initial usage sessions longer than users who pay no such costs.*

Proposition 4. *Those who watch system usage before trying a system will tend to persist longer in initial usage sessions.*

Proposition 5. *Those who show interest in using a system subsequent to system-specific promotion will tend to have previous interests and inclinations related to aspects of that system.*

Proposition 6. *Those who claim interests related to a system are more likely to persist in usage sessions with that system.*

Proposition 7. *Given access to a system, groups embracing activities or values related to a system will tend to produce more users, and more persistent usage, than other groups.*

Proposition 8. *Those who exhibit a greater degree of concentration during system usage will tend to persist in usage longer.*

Proposition 9. *Those who are likely to persist in usage sessions for longer durations are also more likely to vocalize during usage.*

Proposition 10. *Those who are likely to persist in usage sessions for longer durations are also more likely to perform extraneous movement during usage.*

Proposition 11. *Users who learn during usage will tend to persist in usage longer, and/or users who persist in usage longer will tend to learn more.*

Proposition 12. *Users who enjoy social or performative situations, given a system where usage is observable by an audience of strangers, will tend to persist in usage longer than other users.*

Proposition 13. *Those using demonstrative systems in the presence of friends or family will tend to persist in usage longer.*

Proposition 14. *If a system is designed such that people must trade places, this restricts the number of users over time.*

Proposition 15. *If all else is held equal, but the number of people waiting to use a system is increased, the amount of usage time per session will decrease, thus allowing more people to try.*

The design framework in this dissertation is complex; no single evaluation can sufficiently scrutinize all its heuristics. For example, the data in section 5.3 explores heuristics such as promotion, targeting, cognitive richness, and demonstrability more thoroughly than other heuristics such as novelty and explorability. Therefore, in addition to testing the above propositions, further studies need to be undertaken that examine the various heuristics. Since existing research using *pdMusic* focuses more on shorter-term heuristics, longer-term studies need to be undertaken in the future. Systems that are focused on different aspects of the framework need to be created; if enough such systems are studied, eventually the entire framework might be explored in a piecemeal fashion. The following is a list of potential studies and experiments that could be used to further determine the impact of each heuristic:

- Promotion

  - Create a practice-oriented system that is designed for distribution via the web. Offer two identical versions of the system, differing only in name. Promote one more heavily than the other to see the difference in the user base. Based on the data from section 5.3, it would be interesting to determine the difference in how well-targeted system users were.

  - Create a flashy, eye-catching system, then compare it to an identical system that is of a minimalist aesthetic design. Compare usage, paying special attention to the behaviors of those who observe the system before trying it.

- Convenience

  - Create a practice-oriented system that is distributed via internet download. All who navigate to the download page are randomly steered to one of two download sources. One of the two download sources takes twice as long as the other. Count the downloads from each source.

  - Create a practice-oriented system that is distributed via internet download. All who navigate to the download page are randomly steered to one of two download sources. One of the two download sources requires dependencies so that two steps must be taken for installation. The other provides a single package. Count the downloads from each source.

- Targeting

  - Create a practice-oriented system that is based on and similar to an existing hobby. Introduce it to practitioners and non-practitioners of the original hobby, observing the response.

- Novelty

– Create a practice-oriented system that is based on and similar to an existing hobby. Introduce this system to practitioners and non-practitioners of that hobby. Observe how users behave differently between their first and all subsequent sessions. Control for background, as practitioners of the existing hobby may find the interface less novel than non-practitioners.

- Usability

    – Create two similar versions of a practice-oriented interface. In one version, interface usage is occasionally interrupted by a popup survey question that requires answering or closing. Compare the lengths of usage sessions, and the incidence of subsequent sessions, among users of each interface.

- Cognitive Richness

    – Create a practice-oriented system, based on defeating some challenge or puzzle, in two versions. Version A is a simple challenge requiring very little attention. Version B is a similar challenge, but requires careful attention to detail. Observe usage between the two groups, paying special attention to user engagement and usage persistence.

- Incrementality

    – Create a practice-oriented system based on defeating some challenge or puzzle. Design it such that there is a wide range of easy, medium, and hard challenges, such that a gradual increase in difficulty for new users can be engineered. Design it for distribution via the web. Then steer all potential users to one of the following groups: a) a group that receives only easy challenges with no increase in difficulty over time, b) a group that receives no easy challenges and is forced to jump in at a higher level of difficulty, c) a group that starts with easy challenges but then jumps directly to hard challenges with no middle transition, and d) a group that goes through a gradual progression from easy to medium to hard. Observe usage, paying special attention to user affect and usage persistence.

- Immediacy

    – Any experiments designed to understand convenience and usability will also address immediacy.

- Ownership

    – Create a practice-oriented system in two versions, designed for distribution via the web. Randomly steer all new users to one of the two versions. In version A, new users are automatically assigned a generic avatar. In version B, new users are given a variety of configurable options that allow them to customize the appearance and/or behavior of the avatar. Compare usage between the two groups.

    – Create a practice-oriented system, based on some sort of art creation, in two versions. In version A, users are asked to replicate existing art. In version B, users are asked to create their own art. Compare usage between the two groups, making sure to control for user artistic background and familiarity with the interface.

- Explorability

  - Create a practice-oriented system that is based on and similar to an existing hobby that has many long-term adherents. Introduce it to a well-targeted group composed of both practitioners and non-practitioners of the existing hobby. Observe the response, paying careful attention to those who set the system aside and their reasons for doing so.

- Demonstrability

  - Create a practice-oriented system designed for distribution via the web. Ascertain whether system users identify as introvert or extrovert. See if usage is the same or different between groups. Then install the same system into a public event so that an audience is there. Again, ascertain if users identify as introvert or extrovert and look for differences in usage between groups.

  - Create a practice-oriented system that is based on and similar to an existing hobby that has many long-term adherents. Design it so that skill in the original hobby will transfer to the new system. Install the system publicly, then check for correlations between audience size and usage persistence, making sure to factor in user skill. One hypothesis is that audience size and usage persistence will correlate when users are skilled, and inversely correlate when users are unskilled.

- Cooperation

  - Create a practice-oriented system in two versions, where version A is single-player and version B is multi-player. Install at public events and observe for differences in volitional usage.

Some future studies on volitional usage may be based on upcoming versions of *pdMusic*, the next iteration of which is currently under development. In the coming version, the DJ will no longer be necessary for system operation, because she has provided a long mix track for integration directly into the system. Control of system configuration will be returned to the users, who will be able to change settings using only the Kinect sensor (no mouse or keyboard necessary). These changes remove the need for constant supervision of *pdMusic*, affording installation in more locations, and for longer durations. When the system is installed in galleries or showcases, computer-vision techniques will be developed to estimate audience sizes, affording more exact study on the ratio of potential users to actual users. These computer-vision techniques can also be used throughout galleries to determine which exhibitions draw the largest audiences, for the purposes of better understanding heuristics related to promotion,

aesthetics, or cognitive richness. In addition to the gallery installations, a downloadable version of *pdMusic* will be made so Kinect owners can operate the system from home. In my study of the first and second iterations of *pdMusic*, it seems likely that social aspects of the installation environments had great impact on volitional usage. These effects will be controlled for in home usage. Eventually, tools facilitating composition-making—including a track-based note sequencer—will be added to the system, because it is currently challenging to improvise compelling compositions. Also, *pdMusic* will be made into a multiplayer system in order to explore the heuristic of cooperation.

In order to fully test the design framework, however, various divergent systems, that are not like *pdMusic*, must be deployed and tested under different circumstances. For this reason, new systems are being contemplated or are actively under design. Some of these systems are:

Emotional Slideshow Generator: I intend to revisit the emotional slideshow generator (ESG) from section 4.2, making it into an interface for improvised story-making and narration as discussed in section 4.2.2. Unlike *pdMusic*, ESG requires no hardware; this will afford the exploration of web-based delivery methods for enhanced convenience, and will allow some of the web-based experiments listed above to be undertaken. To better understand cooperation and demonstrability, I hope to create a web-based community around the story-improvising activity afforded by this interface.

Hobby-hybrid system: Some heuristics could be explored in greater detail through the creation of systems affording activities that are hybrids of previously-existing hobbies. For example, two existing hobbies are amateur musicianship and yo-yo playing. It would be interesting to see how people respond to an interface combining these activities. This could be accomplished, hypothetically, by controlling music generators with motion sensors built into a yo-yo. It is possible that such hybrid-hobby interfaces would appeal to many participants from both existing hobbies; in other words, the set of people interested in musical yo-yo would be a union

of the sets of existing hobby participants. Alternatively, it is possible that such hobby-hybrid interfaces would appeal mainly to those who already participate in both existing hobbies; i.e., the set of people interested in musical yo-yo would be an intersection of the sets of existing hobby participants. Therefore, finding out the previous interests of a hobby-hybrid interface's user base would inform heuristics such as targeting and ownership. Also, some hobby-hybrid interfaces might provide insight on the topics of learning and incrementality by exploring how skills transfer between activities. For example, in the hypothetical musical yo-yo system described above, manual dexterity is required for each pre-existing hobby, so it would be interesting to see how skill at manual dexterity, acquired through practicing an existing hobby, might transfer to the new interface.

Multiplayer bouncy-ball: In order to target young people who are interested in sports such as basketball, I am contemplating the creation of a two-player bouncy-ball game, where the ball contains a remote-control servo that shifts its weight distribution. One person bounces the ball, while the other person uses the remote control to make the ball bounce and spin in unpredictable ways. This is a form of two-player competition akin to that between the rider of a mechanical bull and the operator of that mechanical bull. This system should help the ball-bouncer build reaction-time and hand speed.

Language puzzles for two people: In an effort to make learning new languages more fun, I am contemplating the creation of a web interface combining video chat and word puzzles designed to facilitate language learning. The puzzles are meant to be solved by two people who speak different languages. Each person can only enter words from the other person's language, so communication (via discussion, pantomime, and drawing) between players must take place in the video chat. Pairs of players will need to be well-matched in terms of their ability with the opposing language. Similar games can be constructed to build skill at pronunciation, phrasing, and conversation.

To conclude this chapter, it is important to note that efforts must continue to make the design framework more concise and workable. For example, when the shorter-term and longer-term sets of heuristics are considered together, the heuristic of immediacy completely overlaps with the heuristics of usability and convenience. For this reason, a reconciliation between shorter-term and longer-term frameworks, such as that given in table 3.3, deserves further consideration in future studies. In the future work discussed in this chapter, I am the common element: all of this work is to be performed by me, so any biases I hold will permeate. One of my fondest hopes is that others will take up scholarship of the topic of volitional usage. Then, future scholars could design their own systems along with methods for studying volitional usage on those systems. With such contributions, the design framework might finally be tested and shaped under a wide enough range of circumstances to be treated as valid and established grounded theory.

Chapter 7

CONCLUSION

This dissertation is focused on the creation of a design framework providing guidance when creating experiential media systems that are meant to draw volitional usage. Relevant sources informing this framework and its creation are discussed in chapter 2 of this document, and the framework, as it currently stands, was introduced in chapter 3. To begin the process of validating its workability, I used it as an analytical tool in chapter 4, examining volitional usage and engagement, in retrospect, on three systems I helped design during my time at Arizona State University. All of these systems were designed to draw volitional usage, at least in shorter terms, and all took different approaches to engaging users, but all had volitional usage problems. These problems inspired aspects of the framework, and the individual approaches taken by these systems allowed me to explore, on an intellectual level, how users react to good or poor implementations of the framework heuristics. The resulting framework explains my recollections of user behavior on these systems, allowing me to speculate on how users might react given systems or circumstances presenting alternate configurations of heuristics.

In order to begin the process of evaluating the framework for fit and modifiability, in chapter 5 I designed a practice-oriented system, *pdMusic*, using the framework as a set of guidelines. I evaluated the system with regard to volitional usage, identifying some design elements needing improvement; this in turn called for an expansion to the framework to include shorter-term volitional usage. This expansion was implemented, and the expanded framework was used to facilitate the design of a second iteration *pdMusic*. At this time, I also identified a challenge inherent to the task of assessing systems for volitional usage; namely, that evaluation methodologies change the experience of users. For this reason, I began to rely more on qualitative forms of assessment such as personal observation and heuristic assessment, as opposed

to quantitative forms of assessment such as usage-logging. Qualitative assessments leave more room for biases to creep in, but can be designed to be less intrusive, therefore less impactful upon subjects' volitional usage.

The second iteration of *pdMusic* was deployed in a number of public events in order to perform an ethnographic study on short-term users of the system. The behaviors observed during this study were explainable by the framework, and the data from the study provides support and insight into many of the framework heuristics. However, some framework heuristics are addressed more thoroughly than others by the data in this evaluation. Since other theoretical frameworks of scope comparable to my design framework, such as the Serious Leisure Perspective (Stebbins, 2007), took decades to explore in their entirety, I do not feel concerned that the entire framework could not be thoroughly evaluated within the course of my doctoral studies. Like many theories based in psychology, my framework might not ever be definitively proven. My experiences designing and analyzing experiential systems supports the framework theories, and every aspect seems logically consistent and plausible, but evidence for the framework currently relies heavily on case study systems. As discussed in chapter 6, future practice-oriented systems and future research into the topic of volitional usage may call for refinements to the framework in order to enhance its workability and fit.

When beginning to undertake this dissertation, my goal was to design systems that were used voluntarily over the course of months or years, like hobbies. In part because of the challenges discussed above, I soon discovered it was difficult enough to design systems that draw usage in the earlier stages. Even the first iteration of *pdMusic*, which was most explicitly designed with long-term engagement in mind, was perceived by most users as engaging for only a few usages. It is worth noting that hobby-like activities tend to have small participant bases. Playing instruments, for example, is one of the most prevalent existing hobbies, but is practiced frequently by only four percent of the population and occasionally by eight percent[1].

---

[1]Per U.S. Census Bureau. See www.census.gov/prod/2011pubs/12statab/arts.pdf

Therefore, it stands to reason that even the most well-designed practice-oriented systems will draw long-term volitional use from only some fraction of potential users. Since the only users with any likelihood of exhibiting long-term volitional usage are those who exhibit shorter-term volitional usage, it became clear that focusing on those early stages—the first one or two usages—was a better course of action for this dissertation.

The main goal of this dissertation was to simplify the task of designers by separating important details, having a large impact on volitional usage, from unimportant details. What does one need to consider when trying to make engaging systems? How can one minimize the length of time spent designing a system while simultaneously maximizing the intensity of engagement and endurability of engagement? How can one balance features of a system so that beginning users enjoy usage, but accustomed users also continue being engaged? After doing the projects in this dissertation, I feel the task of creating systems that draw volitional usage, especially longer-term volitional usage, is still a complex and challenging one. Even knowing the framework heuristics, building these characteristics into systems requires forethought, effort, and iterative development, for the following reasons:

- Tradeoffs abound between the various heuristics, and focusing too much on any single heuristic, to the detriment of other heuristics, may decrease overall volitional usage.

- Each heuristic is broad enough to require a great deal of design consideration in its own right. Determining the meaning of any heuristic in the context of a specific system is a challenging task; and invariably, there exist multiple ways that a heuristic might be implemented.

- Systems incorporating all the heuristics are likely to be highly feature-rich, and these sorts of systems require detailed designs that take time. In some circumstances, designers might reduce this problem by implementing features from which multiple heuristics can emerge.

Still, even though designing practice-oriented interfaces is a complex process, and future research remains to be done on the topic of volitional usage, my design framework can begin to serve its intended function. Its heuristics contain a strong vein of common sense, and can be used as practical design concepts: organizing, guiding, and catalyzing the thought-processes of designers who wish to make practice-oriented systems.

I will close with the following thought. From time to time, technology emerges that is given the label 'disruptive innovation,' due to the way it changes society. Examples of this abound in realms such as personal and mobile computing. In nearly all cases, these new technologies bring good things to society, but occasionally there are negative side effects. For example, the rise of a culture of computing and video gaming in the U.S. has led to great societal advances, but may also be implicated in an abatement of physical exercise, contributing to reduced physical fitness. This was an unintended result, but it is a prime example of interfaces transforming users. The theme of this dissertation can be expressed as follows: designers are needed who understand how usage can beneficially transform users, who design specifically for these beneficial transformations, and who know how to ensure that usage is adopted so the transformations take place. Such designers could address many societal problems that are subject to the willpower of individuals, and the technologies created could make broad impacts upon the public. Thus, through designing for volitional usage, we can begin to intentionally create innovations that disrupt society, and we can ensure that these innovations do so for the better. I hope the ideas in this dissertation will foster insight and inspire readers, such as yourself, to improve the world in this way. I humbly thank you for reading.

REFERENCES

Attfield, S., Kazai, G., Lalmas, M., and Piwowarski, B. (2011). Towards a science of user engagement. In *WSDM Workshop on User Modelling for Web Applications*.

Baldauf, D., Burgard, E., and Wittmann, M. (2009). Time perception as a workload measure in simulated car driving. *Applied Ergonomics*, 40(5):929 – 935.

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, 47(4):644–675.

Bernsen, N. (1994). Foundations of multimodal representations: A taxonomy of representational modalities. *Interacting with Computers*, 6(4):347–371.

Blythe, M., Overbeeke, K., Monk, A., and Wright, P. (2004). *Funology: From Usability to Enjoyment*, volume 3. Springer.

Campana, E., Tanenhaus, M., Allen, J., and Remington, R. (2011). Natural discourse reference generation reduces cognitive load in spoken systems. *Natural Language Engineering*, 17(3):311.

Chen, J., Penhune, V., and Zatorre, R. (2008). Listening to musical rhythms recruits motor regions of the brain. *Cerebral Cortex*, 18(12):2844– 2854.

Chen, Y., Huang, H., Xu, W., Wallis, R., Sundaram, H., Rikakis, T., Ingalls, T., Olson, L., and He, J. (2006). The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation. In *Proceedings of the 14th Annual ACM International Conference on Multimedia*, pages 763–772. ACM.

Chen, Y., Sundaram, H., Rikakis, T., Ingalls, T., Olson, L., and He, J. (2009). Experiential media systems—the biofeedback project. *Multimedia Content Analysis*, pages 1–34.

Collier, G. (2002). Why does music only express some emotions? A test of philosophical theory. *Empirical Studies of the Arts*, 20(1):22–31.

Creswell, J. (2008). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications, Inc.

Csikszentmihalyi, M. (1991). *Flow*. Harper Perennial.

Czerwinski, M., Horvitz, E., and Cutrell, E. (2001). Subjective duration assessment: An implicit probe for software usability. Technical report, Microsoft Research.

Dewey, J. (1934). *Art as Experience*. Minton, Balch.

Dombovy, M. (2004). Understanding stroke recovery and rehabilitation: Current and emerging approaches. *Current Neurology and Neuroscience Reports*, 4(1):31–35.

Duff, M. (2012). *Adaptive Mixed Reality Rehabilitation for Stroke*. PhD thesis, Arizona State University.

Duncan, P., Propst, M., and Nelson, S. (1983). Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Physical Therapy*, 63(10):1606–1610.

Duncan, P., Wallace, D., Lai, S., Johnson, D., Embretson, S., and Laster, L. (1999). The Stroke Impact Scale version 2.0 evaluation of reliability, validity, and sensitivity to change. *Stroke*, 30(10):2131–2140.

Edelson, D. (2002). Design research: What we learn when we engage in design. *The Journal of the Learning Sciences*, 11(1):105–121.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124.

Ekman, P., Levenson, R., and Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*.

Elfenbein, H. A. and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2):203.

Elfenbein, H. A. and Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science*, 12(5):159–164.

Erikson, E. and Erikson, J. (1998). *The Life Cycle Completed*. W.W. Norton & Co.

Farley, B. G. and Koshland, G. F. (2005). Training BIG to move faster: The application of the speed-amplitude relation as a rehabilitation strategy for people with Parkinson's disease. *Experimental Brain Research*, 167(3):462–467.

Freedberg, D. and Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences*, 11(5):197–203.

Fugl-Meyer, A., Jääskö, L., Leyman, I., Olsson, S., Steglind, S., et al. (1975). The post-stroke hemiplegic patient: A method for evaluation of physical performance. *Scandinavian Journal of Rehabilitation Medicine*, 7(1):13.

Gabrielsson, A. and Lindström, E. (2001). The influence of musical structure on emotional expression. In Juslin, P. and Sloboda, J., editors, *Music and Emotion: Theory and Research*, chapter 10. Oxford University Press.

Gaver, W. (1991). Technology affordances. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems: Reaching through Technology*, pages 79–84. ACM.

Gee, J. (2007). *Good video games + good learning: Collected essays on video games, learning, and literacy*, volume 27. Peter Lang Pub Inc.

Gilbert, D. T. (2006). *Stumbling on Happiness*. A.A. Knopf, New York, 1st edition.

Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3):295–314.

Glaser, B. and Strauss, A. (2009). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Transaction Books.

Gundlach, R. (1935). Factors determining the characterization of musical phrases. *American Journal of Psychology*, 47(4):624–643.

Hallam, S. (2002). Musical motivation: Towards a model synthesising the research. *Music Education Research*, 4(2):225–244.

Hassenzahl, M. (2008). User experience (UX): Towards an experiential perspective on product quality. In *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine*, pages 11–15. ACM.

Hassenzahl, M. and Tractinsky, N. (2006). User experience—a research agenda. *Behaviour and Information Technology*, 25(2):91–97.

Hayes, G. (2011). The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(3):15.

He, H. and Jiping, H. (2006). Recent developments in biofeedback for neuromotor rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 3.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.

Jennett, C., Cox, A., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9):641–661.

Johnston, A. (2011). Beyond evaluation: Linking practice and theory in new musical interface design. *Interpretation*, 14:105.

Jordà, S. (2002). FMOL: Toward user-friendly, sophisticated new musical instruments. *Computer Music Journal*, 26(3):23–39.

Juslin, P. and Vastfjall, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(5):559.

Kaptelinin, V. and Nardi, B. (2012). Activity theory in HCI: Fundamentals and reflections. *Synthesis Lectures Human-Centered Informatics*, 5(1):1–105.

Kim, Y., Butzel, J., and Ryan, R. (1998). Interdependence and well-being: A function of culture and relatedness needs. *International Society for the Study of Personal Relationships, Saratoga spring, NY*.

Kjeldskov, J. and Graham, C. (2003). A review of mobile HCI research methods. *Human-Computer Interaction with Mobile Devices and Services*, pages 317–335.

Krueger, M. W. (1977). Responsive environments. In *Proceedings of the National Computer Conference*, pages 423–433. ACM.

Lang, P. J., Bradley, M., and Cuthbert, B. N. (2005). International affective picture system (IAPS): Instruction manual and affective ratings. Technical report, NIMH Center for the Study of Emotion and Attention.

LeCompte, D. (1999). Seven, plus or minus two, is too much to bear: Three (or fewer) is the real magic number. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 43 (3), pages 289–292. SAGE Publications.

Lehrer, N., Attygalle, S., Wolf, S. L., and Rikakis, T. (2011a). Exploring the bases for a mixed reality stroke rehabilitation system, part I: A unified approach for representing action, quantitative evaluation, and interactive feedback. *Journal of Neuroengineering and Rehabilitation*, 8(1):51.

Lehrer, N., Chen, Y., Duff, M., Wolf, S. L., and Rikakis, T. (2011b). Exploring the bases for a mixed reality stroke rehabilitation system, part II: Design of interactive feedback for upper limb rehabilitation. *Journal of Neuroengineering and Rehabilitation*, 8(1):54.

Lehrer, N., Tinapple, D., Koziupa, T., Chen, M., Kidane, A., Rajko, S., Wallis, I., Baran, M., Lorig, D., Siwiak, D., et al. (2010). Thrii. In *Proceedings of the International Conference on Multimedia*, pages 1425–1428. ACM.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. MIT Press.

Leontyev, A. (1981). *The Development of Mind*. Marxists Internet Archive.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. In *Soviet Physics Doklady*, volume 10, page 707.

Liepert, J., Bauder, H., Miltner, W., Taub, E., and Weiller, C. (2000). Treatment-induced cortical reorganization after stroke in humans. *Stroke*, 31(6):1210–1216.

Malone, T. (1982). Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Conference on Human Factors in Computing Systems*, pages 63–68. ACM.

Manturzewska, M. (1990). A biographical study of the life-span development of professional musicians. *Psychology of Music*, 18(2):112–139.

Marshall, S. and Cohen, A. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception*, pages 95–112.

Maslow, A. (1943). A theory of human motivation. *Psychological Review*, 50(4):370.

McCarthy, J. and Wright, P. (2004). Technology as experience. *Interactions*, 11(5):42–43.

Meyer, L. (1956). *Emotion and Meaning in Music*. University of Chicago Press.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81.

Myers, C. and Habiner, L. (1981). A comparative study of several dynamic time-warping algorithms for connected-word recognition. *Bell System Technical Journal*.

Myers, M. (1997). Qualitative research in information systems. *Management Information Systems Quarterly*, 21:241–242.

Niedenthal, P. (2007). Embodying emotion. *Science*, 316(5827):1002–1005.

Nielsen, J. and Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People*, pages 249–256. ACM.

Norman, D. (2002). *The Design of Everyday Things*. Basic books.

Norman, D. (2005). Human-centered design considered harmful. *Interactions*, 12(4):14–19.

Norman, D. (2010). *Living with Complexity*. MIT press.

Obrenovic, Z. and Starcevic, D. (2004). Modeling multimodal human-computer interaction. *Computer*, 37(9):65–72.

O'Brien, H. and Toms, E. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6):938–955.

O'Brien, H. and Toms, E. (2009). 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.

Orme, J. E. (1969). *Time, Experience, and Behavior*. Iliffe Books.

Pacchetti, C., Mancini, F., Agliere, R., Fundaro, C., Martignoni, E., and Nappi, G. (2000). Active music therapy in parkinson's disease: An integrative method for emotional rehabilitation. *Psychosomatic Medicine*, 62(3):386–393.

Persichetti, V. (1961). *Twentieth-Century Harmony: Creative Aspects and Practice*. W.W. Norton & Co.

Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology*, 8(3):179–193.

Rikakis, T., Spanias, A., Sundaram, H., and He, J. (2006). An arts, sciences and engineering education and research initiative for experiential media. In *36th Annual Frontiers in Education Conference*, pages 13–18. IEEE.

Rittel, H. and Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2):155–169.

Rizzo, A., Pair, J., McNerney, P., Eastlund, E., Manson, B., Gratch, J., Hill, R., Swartout, B., et al. (2005). Development of a VR therapy application for Iraq war military personnel with PTSD. *Studies in Health Technology and Informatics*, 111:407–413.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273 – 294.

Ryan, R. and Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68–78.

Sawyer, R. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music*, 34(2):148–165.

Schellenberg, E. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, 58(1):75–125.

Scherer, K. (2000). Psychological models of emotion. *The Neuropsychology of Emotion*, 137:162.

Schwartz, B. (2004). The tyranny of choice. *Scientific American Mind*.

Sengers, P., Boehner, K., Mateas, M., and Gay, G. (2008). The disenchantment of affect. *Personal and Ubiquitous Computing*, 12(5):347–358.

Sengers, P. and Gaver, B. (2006). Staying open to interpretation: engaging multiple meanings in design and evaluation. In *Proceedings of the 6th conference on Designing Interactive Systems*, pages 99–108. ACM.

Skinner, B. (1948). 'Superstition' in the pigeon. *Journal of Experimental Psychology*, 38(2):168.

Sloboda, J. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music*, 19(2):110–120.

Soanes, C. and Stevenson, A., editors (2006). *Concise Oxford English Dictionary*. Oxford University Press.

Stebbins, R. (1982). Serious leisure: A conceptual statement. *Pacific Sociological Review*, pages 251–272.

Stebbins, R. (2007). *Serious leisure: A perspective for our time*. Transaction Publishers.

Sundaram, H. and Rikakis, T. (2006). Experiential media systems. *Encyclopedia of Multimedia*, pages 212–217.

Swartz, L. (2003). Why people hate the paperclip: Labels, appearance, behavior and social responses to user interface agents. Honors Thesis, Stanford.

Swift, B., Gardener, H., and Riddell, A. (2013). Chasing a feeling: Experience in computer support jamming. In Holland, S., Wilkie, K., Mulholland, P., and Seago, A., editors, *Music and Human-Computer Interaction*, chapter 5. Springer-Verlag.

Tinto, V. (1987). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press.

Turkle, S. (2008). *Falling for Science: Objects in Mind*. MIT Press.

Uswatte, G., Taub, E., Morris, D., Vignolo, M., and McCulloch, K. (2005). Reliability and validity of the upper-extremity Motor Activity Log-14 for measuring real-world arm use. *Stroke*, 36(11):2493–2496.

Van den Akker, J. (1999). Principles and methods of development research. *Design Methodology and Developmental Research in Education and Training*, pages 1–14.

Vermeire, E., Hearnshaw, H., Van Royen, P., and Denekens, J. (2001). Patient adherence to treatment: Three decades of research. A comprehensive review. *Journal of Clinical Pharmacy and Therapeutics*, 26(5):331–342.

Vuong, C. H., Ingalls, T., and Abbas, J. J. (2011). Transforming clinical rehabilitation into interactive multimedia. In *Proceedings of the 19th ACM International Conference on Multimedia*, pages 937–940. ACM.

Wallis, I., Ingalls, T., and Campana, E. (2008). Computer-generating emotional music: The design of an affective music algorithm. In *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espöo, Finland.

Wallis, I., Ingalls, T., Campana, E., and Goodman, J. (2011a). A rule-based generative music system controlled by desired valence and arousal. In *Proceedings of the 8th International Sound and Music Computing Conference (SMC)*, Padova, Italy.

Wallis, I., Ingalls, T., Campana, E., and Vuong, C. (2011b). The instrumental hobby and long-term engagement with HCI. In *BCS HCI 2011 Workshop—When Words Fail: What can Music Interaction tell us about HCI?*, Newcastle upon Tyne, UK.

Wallis, I., Ingalls, T., Campana, E., and Vuong, C. (2013). Amateur musicians, long-term engagement, and HCI. In Holland, S., Wilkie, K., Mulholland, P., and Seago, A., editors, *Music and Human-Computer Interaction*, chapter 3, pages 49–66. Springer-Verlag.

Wallis, I., Ingalls, T., Rikakis, T., Olsen, L., Chen, Y., Xu, W., and Sundaram, H. (2007). Real-time sonification of movement for an immersive stroke rehabilitation environment. In *Proceedings of the 13th International Conference on Auditory Display (ICAD)*, Montréal, Canada.

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

Whitson, J. and Galinsky, A. (2008). Lacking control increases illusory pattern perception. *Science*, 322(5898):115–117.

Wolf, S., Catlin, P., Ellis, M., Archer, A., Morgan, B., and Piacentino, A. (2001). Assessing Wolf motor function test as outcome measure for research in patients after stroke. *Stroke*, 32(7):1635–1639.

Xu, W., Chen, Y., Sundaram, H., and Rikakis, T. (2006). Multimodal archiving, real-time annotation and information visualization in a biofeedback system for stroke patient rehabilitation. In *Proceedings of the 3rd ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, pages 3–12. ACM.

Zentner, M., Didier, G., and Scherer, K. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4):494–521.

Zimmerman, J., Forlizzi, J., and Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 493–502. ACM.

Zimmerman, J., Stolterman, E., and Forlizzi, J. (2010). An analysis and critique of research through design: Towards a formalization of a research approach. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, pages 310–319. ACM.

BACKGROUND SONGS AND KEYS IN *PDMUSIC*

Title	Artist	Key
Needin You	Sek	D♭ minor
Smoke in the Club	Peter Horrevorts	B minor
Into a Deep	Alain Ho	A minor
Change in You	Jimpster	B minor
Look Straight	The Private Lightning	B minor
Into the Wild	JML	C minor
Kerpow	XXXY	D minor
I'm Deep I'm Soul	Nuyorica	E minor
Masks	Martyn	D minor
Stoer	Ruben de Ronde	B minor
Someone is Out	Gabriel Garcia	C minor
Margeaux	Kahn	E♭ minor
MD2.4.2	MD2	B♭ minor
Save Love	Wally Lopez	D minor
The Lord's Graffitti	Actress	A♭ minor
Cream	Dirty Mongrel	G minor
I Feel For You	Azuni	G♭ minor
Let's Go Juno	Rejected	C minor
Ushauia	Motorcitysoul	A♭ minor
I'm Com'in	DJ JUS ED	B♭ minor
Sleep in Tokyo	SBTRKT	C minor
ABAW 723 (Original Mix)	Phon.o	E minor
Grow (Original Mix)	Roman Lindau	B minor
Goldene Kugel (Original Mix)	Siriusmo	A♭ major
Bass Comedown (Original Mix)	DJ Madd	A minor
Sexy Motherfucker (Original Mix)	DJ Martjin Van Dishoeck, DJ Wiebthroat	F minor
Turbosteppa	Claude VonStroke & Kry Wolf	G♭ minor
John Doe	Detroit Swindle	F minor
Go On Then (feat. Jem Cooke, Original Mix)	Kate Simko	G minor
Love is a Lie (Dale Howard Remix)	Luis Leon	G minor
Howling (Âme Remix)	Ry & Frank Wiedemann	G♭ minor
Hardbody	Scuba	A minor
Jammin' (Huxley Remix)	SHOW-B	E minor

Table A.1: Songs and musical keys of background music (provided by the *pdMusic* DJ) used in second iteration of *pdMusic*.

APPENDIX B

SURVEY QUESTIONS IN *PDMUSIC* PILOT STUDY

B.1 Introductory Questions

- Do you consider yourself a music-maker of any sort? If so, in what way do you make music? Approximately how many hours per week do you spend making music? Please write your answer in the space below.

- Do you practice any movement arts (i.e. martial arts, dancing, etc.)? If so, what movement art do you practice? Approximately how many hours per week do you spend practicing movement arts? Please write your answers in the space below.

- Do you enjoy video games? Can you name any video game titles or genres you have found particularly enjoyable? Approximately how many hours per per week do you spend playing video games? Please write your answers in the space below.

- Do you have a hobby or passion which you do frequently? If so, what is it? Please note, for this question we are interested only in hobbies which require some type of skill. This rules out many forms of entertainment such as tv-watching or reading. Please write your answer in the space below.

- Are you male or female? Please write your answer in the space below.

- How old are you? Please write your answer in the space below.

- Where did you grow up? Please write your answer in the space below.

B.2 Engagement Questions

- Video games are often rated according to the hours of gameplay they deliver. In your judgement, how many hours could *pdMusic* deliver?

- Using this 1 to 5 scale, please rate how fun *pdMusic* is to you. 1 is not fun, and 5 is highly fun. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate the addictiveness of *pdMusic*? 1 is not addictive and 5 is highly addictive.

- Using this 1 to 5 scale, can you rate your own mastery of *pdMusic*? 1 means you consider yourself a novice, and 5 means you consider yourself a master.

- Using this 1 to 5 scale, how skilled would you say you currently are at *pdMusic*? 1 means you do not feel skilled, and 5 means you feel highly skilled.

- Using this 1 to 5 scale, can you rate the overall difficulty of *pdMusic*? 1 means easy, 5 means difficult. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate how quickly you believe you are attaining skill at *pdMusic*? 1 means you feel you are getting no better, and 5 means you feel you are advancing rapidly.

- Using this 1 to 5 scale, can you rate the startup and launch of *pdMusic* in terms of ease and quickness? 1 means not quick or easy, and 5 means highly quick and easy. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate how *pdMusic*'s location in the student lounge affects your enjoyment? 1 means you enjoy it less because of its location, and 5 means you enjoy it more. Please elaborate in the space below.

- Using this 1 to 5 scale, could you rate the difference in how frequently you would play *pdMusic* if you could carry it from place to place and play it anywhere? 1 means you would play it much less, and 5 means you would play it much more. Please elaborate in the space below.

- Using this 1 to 5 scale, could you rate the difference in how frequently you would play *pdMusic* if it were installed in your home? 1 means you would play it much less, and 5 means you would play it much more. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate the *pdMusic* composition (created by someone else) you most recently played in terms of challenge? 1 means easy, 5 means hard, and 3 means challenging but doable. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate the *pdMusic* composition you most recently created in terms of playing challenge? 1 means easy, 5 means hard, and 3 means challenging but doable. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate how the ability to create and share games impacts your experience? 1 means this ability negatively impacts your experience, and 5 means this ability positively impacts your experience. Please elaborate in the space below.

- Using this 1 to 5 scale, please rate your preference for playing existing *pdMusic* compositions verses creating new *pdMusic* compositions. 1 means you heavily prefer playing existing compositions, and 5 means you heavily prefer making new compositions. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate your ability to develop unique strategies or styles of play in *pdMusic*? 1 means there is no ability to develop unique strategies or styles of play, and 5 means it is very easy to develop unique strategies or styles of play. Please elaborate in the space below.

- How do you feel about your own style of play in *pdMusic*? Would you say your playing style is very unique or very typical? Would you say it is easy or hard for other people to emulate? Please write your answers in the space below.

- Using this 1 to 5 scale, can you rate your own variety of movement when playing *pdMusic*? 1 means your movement tends to be very repetitious, and 5 means it tends to vary widely. Please elaborate in the space below.

- Using this 1 to 5 scale, how would you rate your freedom of movement in *pdMusic*? 1 means *pdMusic* is restrictive, and 5 means it is free. Please elaborate in the space below.

- Using this 1 to 5 scale, please rate how engaging you think *pdMusic* might be if it were a multi-player game. 1 means you think a multi-player version would be much less engaging, and 5 means you think it could be much more engaging. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate the extent to which the ability to create and share *pdMusic* compositions adds or detracts from your enjoyment? 1 means this ability makes *pdMusic* less enjoyable, where 5 means this ability makes *pdMusic* more enjoyable. Please elaborate in the space below.

- Do you compare your scores with other *pdMusic* players as a form of competition? If so, please describe how in the space below.

- Do you compare your *pdMusic* compositions with those of other *pdMusic* players as a form of competition? If so, please describe how in the space below.

- Have you played *pdMusic* along with friends? If so, can you rate on this 1 to 5 scale how this impacted your enjoyment? 1 means having friends along made *pdMusic* much less enjoyable, and 5 means having them made *pdMusic* much more enjoyable. Please elaborate in the space below.

- Have you played *pdMusic* in competition with other players? If so, can you rate on this 1 to 5 scale whether competition makes *pdMusic* more or less enjoyable? 1 means competition makes *pdMusic* much less enjoyable, and 5 means competition makes it much more enjoyable. Please describe in the space below.

- Using this 1 to 5 scale, can you rate how competitive *pdMusic* is? 1 means it is not competitive at all, and 5 means it is highly competitive. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate how collaborative *pdMusic* is? 1 means it is not collaborative, and 5 means it is highly collaborative. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate how social *pdMusic* is? 1 means it is not social, and 5 means it is highly social. Please elaborate in the space below.

- Using this 1 to 5 scale, can you rate the impact of onlookers in the room on your enjoyment of *pdMusic*? 1 means onlookers make playing *pdMusic* much less enjoyable, and 5 means they make it much more enjoyable. Please elaborate in the space below.

- Do you notice any design features which could be improved about *pdMusic*? Please describe in the space below.

- Do you play *pdMusic* more or less frequently than you used to? In the space below, could you describe the reason for this change?

APPENDIX C

INSTITUTIONAL REVIEW BOARD DOCUMENTS

C.1   IRB application for first iteration of *pdMusic*.

| ASU | Arizona State University
Office of Research Integrity
and Assurance
660 S. Mill Avenue Suite 315
Arizona State University
Tempe AZ 85287-6111
(Mail Code 6111)
Phone: 480-965-6788
Fax: (480) 965-7772 | ASU | *For Office Use Only:*
Date Received:
HS Number: |

SOCIAL BEHAVIORAL APPLICATION HUMAN SUBJECTS

PROTOCOL INFORMATION

Protocol Title: Date:
Long-term Engagement with a Computer Interface   Nov 7, 2011

PRINCIPAL INVESTIGATOR (PI)

Please note that the PI's CV and human subject's protection training certification must be attached with this application.

Name and Degree(s):
      Todd Ingalls, MM

Department/Center:
      School of Arts, Media, and Engineering

Mailing Address:
      PO Box 878709, Tempe, AZ 85287

Email: Phone: Fax:
      Todd.Ingalls@asu.edu   (480)455-7590   (480)964-0961

University Affiliation:
  Associate Professor

CO-INVESTIGATORS (CO-I)

- A Co-I is anyone who has responsibility for the project's design, implementation, data collection, data analysis, or who has contact with study participants.
- If the project involves medical procedures or patient care that the PI is not certified or licensed to conduct, a responsible physician or other certified or licensed professional must be included as a Co-I. The application must
  include a copy of supporting documentation for this individual (CV, license, board certification etc).

Name	Study Role	Affiliation	Department	Email/Tel/Fax	Student (yes/no)
Todd Ingalls	Advisory	Associate Professor	AME	Todd.Ingalls@asu.edu	no
Isaac Wallis	System Design	Grad Student	AME	iwallis@asu.edu	yes

Social Behavioral IRB Application Form – Page   1
Revised April 2011

PROJECT FUNDING

1a)How is the research project funded? (A copy of the grant application **must** be provided prior to IRB approval)
Research is not **funded**

b) What is the source of funding or potential funding? (Check all that apply)
Federal   Private Foundation   Department Funds
Subcontract   Fellowship   Other

c) Please list the name(s) of the sponsor(s):

d) What is the grant number and title?

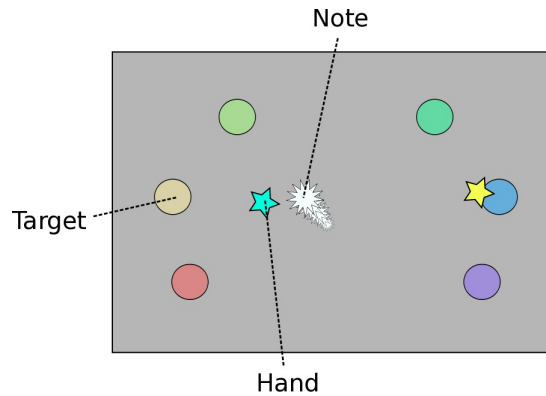e) What is the ASU account number/project number?

f) Identify the institution(s) administering the grant(s):

PROJECT SUMMARY

2. Provide a **brief** description of the **background, purpose, and design** of your research. Avoid using technical terms and jargon. Describe all interactions with potential study participants (e.g., how identified, how recruited) including all of the **means you will use to collect data** (e.g. instruments, measures, tests, questionnaires, surveys, interview schedules, focus group questions, observations). Provide a short description of the tests, instruments, or measures. (If you need more than a few paragraphs, please attach additional sheets.) **Attach copies of all instruments and questionnaires. FOR ALL OF THE QUESTIONS, WRITE YOUR ANSWERS ON THE APPLICATION RATHER THAN SAYING "SEE ATTACHED".**

We have created a set of theories on how to design interactive media which people will find engaging over a long period of time. If these theories are correct, this gives us a framework for designing interfaces to help people rehabilitate or achieve other long-term goals requiring dedication and persistence. To test these theories, we used them to design a musical video game. This video game uses the Microsoft Kinect game controller. Players use arm movements to play melodies along with a background song, and their performance is rated by the game. This is a common gaming paradigm, with a well-known corollary in Guitar Hero; however, certain design factors (such as the ability for players to create their own melodies and share them) should result in heightened long-term engagement. To test this, we want to install this as an arcade game into Stauffer B, the digital culture building. Anyone who wishes may create a login and password for this game and then play it freely whenever they wish. Real names will not be required to register for a username (although email addresses will be required). The login accounts will be used to track the frequency with which people return and play the game; this data will be used to gauge the overall engagement inherent in the system over time. Data will also be collected in the form of occasional survey questions (designed to ascertain the engagement levels of participants and to gauge the effectiveness of certain factors of the game's design). Features of the game (such as the ability to create melodies) will be disabled for certain accounts, so we can tell if these features add or remove from the long-term engagement inherent in the game. The game will remain installed in Stauffer for approximately 3 months.

GAME DESCRIPTION:

2

Note

Target

Hand

The video game uses a visual interface with layout similar to that seen in Figure 2 (This is not a screenshot, it just shows how things are laid out). This interface contains three interactive elements: 1) *Hand Cursors* are visual indicators, the locations of which are driven by motion tracking on the player's hands—we use Microsoft's Kinect sensor to perform this hand tracking. 2) *Targets* are areas of the virtual space that the player must touch in accordance with the game's sequence of actions. Each target is attached to a musical pitch. 3) *Note Indicators* make up the sequence of actions; these radiate out from the centre toward individual targets. Whenever a note indicator reaches a target, a hand cursor must simultaneously touch that target. If this happens, the target's musical tone, which is a melodic tone designed to accompany the background track, will play. If not, a less pleasant "flubbed" sound will occur. Scoring is based on how well the players can match the note indicators by touching the targets.

In addition, some users will have access to a creative mode which allows them to create their own melodies, which others can then play. They will improvising these melodies by touching the targets, and their actions will be recorded so that other students can play that game later: the note indicators will go to the same targets (at the same timings) as were touched by the original game creator. It is our hypothesis that users with access to the creative mode will be more engaged than users without access to the creative mode.

STUDY DURATION
3. What is the expected duration of the study through data analysis? (Include a timeline, if applicable).
1 Year
a. When is the expected date that you wish to begin research? (MM/DD/YY)   01/05/12   (must be after submission date) Note: Protocols are approved for a maximum of 1 year. If a project is intended to last beyond the approval period, continuing review and reapproval are necessary. Research cannot begin until you have received an approval letter.

IRB APPROVAL
4. Has this project been reviewed by another IRB?
No
a) What is the name of the institution?

b) What is the current IRB approval date/status of IRB application?

STUDY SITES
5. Where will the study be conducted? (Check all that apply)
 On campus. (Please indicate building(s) and room number (s) when known)
In building Stauffer B. Room unknown at this time.

6a) What is the expected number of individuals to be <u>screened</u> for enrollment?
 Any number of individuals could participate, and there is no screening—anyone who chooses may participate.

b)What is the **<u>MAXIMUM</u>** number of subjects that you plan to enroll in the study?
It will be surprising if more than 100 choose to participate.

c) What is the approximate number of:
Males   Unknown
Females   Unknown

d) Indicate the age range of the participants that you plan to enroll in your study.   18   to   100
Most will be university undergraduate and graduate students, although others may choose to participate as well. The exact age range is unknown.

e) What is the expected duration of participation for each subject? (at each contact session and total)
Approximately 3 months. During this time, the game will be set up and available for each participant to play. We will track the frequency and time spent by each account holder who plays the game, and this data will be used to judge how engaging the game is. If we notice that an account holder has lost interest with the game, we will send some questions about their experience via email.

SUBJECTS

7.   Will the study involve any of the following participants?
No. (Please check all that apply if your study specifically targets these populations)
 Children (under 18)   Pregnant women
 Prisoners or detainees   Persons at high risk of becoming detained or imprisoned
 Decisionally impaired   Patients- what is the status of their health?
 Fetuses   Native Americans
 Non-English speakers (Include copy of all materials in language of participants and certification of the translation and back-translation: http://researchintegrity.asu.edu/humans/forms)

a) If **any** of the above categories have been checked, please state how you will protect the rights and privacy of these individuals.
None checked.

b) Please provide the rationale for the choice of the subjects including any inclusion criteria.
This is a test of how engaging the interface is. Users freely choosing to participate are essentially self-selecting based on interest, and the test is testing how long we can keep them interested.

c) Will any ethnic/racial or gender groups be excluded from this study? If so, provide the rationale for the exclusion criteria.
No exclusions.

RECRUITMENT

8. Describe the process(es) you will use to **recruit participants** and inform them about their role in the study.
Participants will have access to a video game which is being evaluated. It will be installed in a well-travelled area. Passers-by will be able to play this game for free after creating a login and password for it. At the time of login creation, their role in the study will be displayed for them to read. Participants will be made aware of the system's existence, and the study will be described, during in-class demonstrations. There will also be a cover letter prominently displayed at the site of the game installation and a stack of cover letters which participants can freely take away.

a) Will any of the following be used? (Check all that apply and attach copies)
 Internet/Email

4

Newspapers/radio/televisi on advertising

Posters/brochures/letters Cover Letter (attached)

Other Demonstrations of the interface will be performed at specific times, most likely during classes in Stauffer B, where the interface will be installed for evaluation. Students will be made aware of the interface's existence, and it will be expressed that they have free access to the interface, but that their participation will be both voluntary and anonymous. They will be placed under no obligation, as any obligation to participate would distort the results of the study.

b) Does any member of the research team have a relationship (i.e., teacher, coach, physician, therapist, service provider, etc) with individuals who will be recruited for this study or with institutions that will be used to recruit for this study? If yes, describe this relationship in detail and explain how the research process will avoid any potential problems (e .g, coercion or appearance of possible coercion in recruiting) or conflicts of interest arising from this investigator's dual roles.
The area the game will be installed in is an area where Digital Culture students take classes. The study researchers teach and TA in the digital culture department. However, the study is designed such that participants should be anonymous to the researchers, and this will be known to participants beforehand.

DECEPTION
9.   Does the proposed research require that you deceive participants in any way?
No

a) If your response is "yes," describe the type of **deception** you will use, indicate why it is necessary for this study, and provide a copy of the debriefing script.

COMPENSATION
10. Will any type of compensation be used? (e.g. money, gift, raffle, extra credit, etc)
a)
No (This is a study on engagement. Compensation would distort the results.)

b) Explain why the compensation is reasonable in relation to the experiences of and burden on participants.

c) Is compensation for participation in a study or completion of the study? (Note: participants must be free to quit at any time without penalty including loss of benefits).
 Participation   Completion

d) If any of the participants are economically disadvantaged, describe the manner of compensation and explain why it is fair and not coercive.

INFORMED CONSENT
11. Describe the procedures you will use to **obtain and document informed consent and assent**. **Attach copies of the forms that you will use.** In the case of secondary data, please attach original informed consent or describe below why it has not been included. Fully justify a request for a waiver of written consent or parental consent for minors.
(The ASU IRB website has additional information and sample consent and assent forms.)
As this study uses anonymous participation, we will use cover letters instead of signed consent forms. A cover letter will be prominently posted at the site of installation, and a stack of cover letters will be available there for participants to take home. The cover letter we are using is attached.

RISKS
12. What are the potential risks of the research? (Check all that apply)
No apparent physical, social, emotional, privacy, criminal, employability, or legal risks.

a) Describe any potential risks to human subjects and the steps that will be taken to reduce the risks. Include any risks to the subject's well-being, privacy, emotions, employability, criminal, and legal status. Risks are minimal. Similar to the risk of playing a Microsoft Kinect video game at home. There is no social stigma against playing video games. Survey questions provided to participants will be designed to ascertain their engagement levels or get their impressions on the system design, and will not be sensitive or personal in nature.

BENEFITS

13a) What are the potential benefits to the individual subject, if any, as a result of being in the study? This is a study of engagement, and we hope participants will find our interface engaging and enjoyable.

b) What are the potential benefits, if any, to others from the study? This study will help us better understand how to design technologies that are engaging to people over a long duration. Such technologies will be helpful in many applications related to achieving long-term goals or overcoming challenges. We intend to eventually use this framework to design long-term engaging interfaces for helping people rehabilitate injuries and disorders.

DATA USE

14. How will the data be used? (Check all that apply)
Dissertation, Thesis, Publication/journal article, Conferences/presentations

PROTECTION OF CONFIDENTIALITY

15. Describe the steps you will take to ensure the confidentiality of the participants and data.

a) Indicate how you will safeguard data that includes identifying or potentially identifying information (e.g. coding). Participants who choose to create a login to our arcade game need never provide their names. The only potential identifying information we will have access to is: 1) login names, which are chosen by participants (but if these are poorly chosen they may give clues to the participant's identity), and 2) participant email addresses, which will be used only in the event of a forgotten username or password. The login accounts will be used only to track the frequency with which people return to play the game as a way to measure engagement. The identities of these people will not be used in data analysis however.

b) Indicate when identifiers will be separated or removed from the data.
At the end of data collection, the emails, logins, and passwords of participants will be deleted, and all recorded data for each participant will be associated with a randomly-selected numerical ID.

c) Will the study have a master list linking participants' identifying information with study ID codes, and thereby, their data? If so, provide a justification for having a master list. (Note: In many cases, the existence of a master list is the only part of a study that raises it above minimal risk, that is, places participants at risk.) No. User logins and passwords will be linked to email addresses in order to allow participants to recover forgotten passwords. We may also communicate questions with participants via email, etc. Otherwise the logins exist only to allow us to track the frequency with which participants play the game.

d)If you have a master list and/or data with identifiers, where on campus will the list and/or data be kept? The game will be run using an ASU computer in an ASU space. **(Data sets with identifiers and master lists, whether electronic or in hard copy, should be securely stored on an ASU campus except in unusual circumstances (e.g., research conducted out of the state or country).)**

e) If you have a master list, when will it be destroyed?
We have no master list. See question 15c.

6

f) How long do you plan to retain the data?
1 year.

g) How will you dispose of the data?
The data is in electronic form. All backups of the data will be deleted.

h) Where on campus will you store the signed consent, assent, and parental permission forms
We are going to be using cover letters instead of signed consent forms.
(If applicable)? (**Consent, assent, and parent permission forms should be securely stored on an ASU campus**)

INVESTIGATOR INTERESTS

16Have all investigator filed a current annual conflict of interest questionnaire with the ASU Office of Research Integrity and Assurance? It is the COEUS module at: http://researchintegrity.asu.edu/coi
Yes

a) Do any of the researchers or their family members, have a financial interest in a business which owns a technology to be studied and/or is sponsoring the research?
No

b) Are there any plans for commercial development related to the findings of this study?
No

c) Will the investigator or a member of the investigator's family financially benefit if the findings are commercialized?
No

d) Will participants financially benefit if the findings are commercialized?
No

BIOLOGICAL MATERIALS

17a) Will biological materials be collected from subjects or given to subjects?
No (If no, please skip to question 18)

TRAINING

18. The research team must document completion of human subjects training from within the past 3 years.
(**For more information see:** http://researchintegrity.asu.edu/training/humans)

Please provide the date that the PI and co-investigators completed the training and attach the certificate.
Isaac Wallis – 11/01/11
Todd Ingalls - 12/11/11

PRINCIPAL INVESTIGATOR

In making this application, I certify that I have read and understand the ASU Procedures for the Review of Human Subjects Research and that I intend to comply with the letter and spirit of the University Policy. Changes in to the study will be submitted to the IRB for written approval prior to these changes being put into practice. **I also agree and understand that informed consent/assent records of the participants will be kept for at least three (3)**

years after the completion of the research. **Attach a copy of the PI's CV unless one is already on file with the Office of Research Integrity and Assurance.**

Name (first, middle initial, last):

Signature: _____ Date: 12-14-11

FOR OFFICE USE:	This application has been reviewed by the Arizona State University IRB:
	Full Board Review
	Expedite Categories:
	Exempt Categories:
	Approved Deferred Disapproved
	Project requires review more often than annual Every months
	Signature of IRB Chair/Member: Date:

8

C.2 IRB amendment for second iteration of *pdMusic*.

<table>
<tr><td>
ASU

Arizona State University
Office of Research Integrity
and Assurance
660 S. Mill Avenue Suite 315
Arizona State University
Tempe AZ 85287-6111
(Mail Code 6111)
Phone: 480-965-6788
Fax: (480) 965-7772
</td><td>
ASU
</td><td>
For Office Use Only:
Date Received:
HS Number:
</td></tr>
</table>

SOCIAL BEHAVIORAL APPLICATION HUMAN SUBJECTS

PROTOCOL INFORMATION

Protocol Title: Date:
Engagement with a Musical Interface at Public Events   Nov 7, 2011

PRINCIPAL INVESTIGATOR (PI)

Please note that the PI's CV and human subject's protection training certification must be attached with this application.

Name and Degree(s):
    Todd Ingalls, MM

Department/Center:
    School of Arts, Media, and Engineering

Mailing Address:
    PO Box 878709, Tempe, AZ 85287

Email: Phone: Fax:
    Todd.Ingalls@asu.edu (480)455-7590   (480)964-0961

University Affiliation:
 Associate Professor

CO-INVESTIGATORS (CO-I)

- A Co-I is anyone who has responsibility for the project's design, implementation, data collection, data analysis, or who has contact with study participants.
- If the project involves medical procedures or patient care that the PI is not certified or licensed to conduct, a responsible physician or other certified or licensed professional must be included as a Co-I. The application must
include a copy of supporting documentation for this individual (CV, license, board certification etc).

Name	Study Role	Affiliation	Department	Email/Tel/Fax	Student (yes/no)
Todd Ingalls	Advisory	Associate Professor	AME	Todd.Ingalls@asu.edu	no
Isaac Wallis	System Design	Grad Student	AME	iwallis@asu.edu	yes

Social Behavioral IRB Application Form – Page   1
Revised April 2011

253

1a)How is the research project funded? (A copy of the grant application **must** be provided prior to IRB approval)
 Research is not **funded**

b) What is the source of funding or potential funding? (Check all that apply)
 Federal   Private Foundation   Department Funds
 Subcontract   Fellowship   Other

c) Please list the name(s) of the sponsor(s):

d) What is the grant number and title?

e) What is the ASU account number/project number?

f) Identify the institution(s) administering the grant(s):
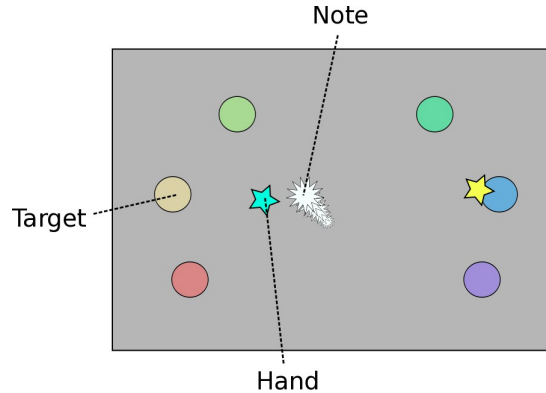
PROJECT SUMMARY

2. Provide a **brief** description of the **background, purpose, and design** of your research. Avoid using technical terms and jargon. Describe all interactions with potential study participants (e.g., how identified, how recruited) including all of the **means you will use to collect data** (e.g. instruments, measures, tests, questionnaires, surveys, interview schedules, focus group questions, observations). Provide a short description of the tests, instruments, or measures. (If you need more than a few paragraphs, please attach additional sheets.) **Attach copies of all instruments and questionnaires. FOR ALL OF THE QUESTIONS, WRITE YOUR ANSWERS ON THE APPLICATION RATHER THAN SAYING "SEE ATTACHED".**

We have created a set of theories on how to design interactive media which people will find engaging over a long period of time. If these theories are correct, this gives us a framework for designing interfaces to help people rehabilitate or achieve other long-term goals requiring dedication and persistence. To test these theories, we used them to design a musical video game. This video game uses the Microsoft Kinect game controller. Players use arm movements to play melodies along with a background song, and their performance is rated by the game. This is a common gaming paradigm, with a well-known corollary in Guitar Hero; however, certain design factors (such as the ability for players to create their own melodies and share them) should result in heightened long-term engagement. We have also adapted this system for use as a game for parties and public events, by creating an interface for electronic music DJs to use, where members of the audience can use the system to improvise music along with the DJ and the DJ interacts with the users by changing the sounds and triggering the rhythm games for them.

To test this system, we will install it during one-day public events held around the Phoenix metro area. Anyone attending these events may play with the system while a DJ plays. I will be taking an observational role in these events, taking notes on how participants and audiences react to the system. I will not be taking note of people's names or any identifying characteristics. If I interact with users at all, it will be to instruct them on how to use the system. There is only one known question I hope to ask of users after these experiences, which is: "If this system were installed at another event you were attending, do you think you would try it again?" Other notes I take will be essentially ethnographic observations, having to do with things such as facial expressions (smiling, confusion), vocalizations, verbalizations, dancing, and so forth. Since it is difficult to take written notes while interacting with users and the DJ, I would also like to record the sound of my system being used—this allows me to take voice notes and hear any statements that pertain to engagement that system users might make. The use of audio recording will be disclosed to each system user before usage. Names will not be asked for during the recording, but if a name is accidentally recorded, (for example, if someone in the background calls a user by name) then

2

that portion of the audio recording will be deleted on the first pass through the data. If users object to being recorded, or if the user is not age 18 (this is unlikely) audio recording will not be performed.

SYSTEM DESCRIPTION:

The video game uses a visual interface with layout similar to that seen in Figure 2 (This is not a screenshot, it just shows how things are laid out). This interface contains three interactive elements: 1) *Hand Cursors* are visual indicators, the locations of which are driven by motion tracking on the player's hands—we use Microsoft's Kinect sensor to perform this hand tracking. 2) *Targets* are areas of the virtual space that the player must touch in accordance with the game's sequence of actions. Each target is attached to a musical pitch. 3) *Note Indicators* make up the sequence of actions; these radiate out from the centre toward individual targets. Whenever a note indicator reaches a target, a hand cursor must simultaneously touch that target. If this happens, the target's musical tone, which is a melodic tone designed to accompany the background track, will play. If not, a less pleasant "flubbed" sound will occur. Scoring is based on how well the players can match the note indicators by touching the targets.

In addition, some users will have access to a creative mode which allows them to create their own melodies, which others can then play. They will improvise these melodies by touching the targets, and their actions will be recorded so that other students can play that game later: the note indicators will go to the same targets (at the same timings) as were touched by the original game creator. It is our hypothesis that users with access to the creative mode will be more engaged than users without access to the creative mode.

STUDY DURATION

3. What is the expected duration of the study through data analysis? (Include a timeline, if applicable).
1 Year
a. When is the expected date that you wish to begin research? (MM/DD/YY)   10/12/12   (must be after submission date) Note: Protocols are approved for a maximum of 1 year. If a project is intended to last beyond the approval period, continuing review and reapproval are necessary. Research cannot begin until you have received an approval letter.

IRB APPROVAL

4. Has this project been reviewed by another IRB?
This digital system is part of an IRB, protocol # 1112007163, which was exempt. This new IRB discusses a slightly modified system (the old system had a login system and a survey questionnaire which are disabled in this system, and this system is changed to work along with a DJ playing music), a new location (art venues around the area), and a new way of recruiting (letting people use the system as they wish at public events)
a) What is the name of the institution?   ASU
b) What is the current IRB approval date/status of IRB application?   Exempt

STUDY SITES

5. Where will the study be conducted? (Check all that apply)
At public events in the Phoenix area. Known locations are: Mesa Center for the Arts, the Ice House (An arts venue).
(Please indicate building(s) and room number (s) when known)

SAMPLE SIZE/DURATION

6a) What is the expected number of individuals to be screened for enrollment?
Any number of individuals could participate, and there is no screening—anyone who chooses may participate.

b)What is the **MAXIMUM** number of subjects that you plan to enroll in the study?
It will be surprising if more than 100 choose to participate.

c) What is the approximate number of:
Males   Unknown
Females   Unknown

d) Indicate the age range of the participants that you plan to enroll in your study.   Unknown

e) What is the expected duration of participation for each subject? (at each contact session and total)
Chosen by the user. In most cases, a few minutes, but if the participant is highly engaged and wishes to continue, then they can. It will be surprising if anyone chooses to use the system longer than twenty minutes at once.

SUBJECTS

7.   Will the study involve any of the following participants?
. (Please check all that apply if your study specifically targets these populations)
Children (under 18)   Pregnant women
Prisoners or detainees   Persons at high risk of becoming detained or imprisoned
Decisionally impaired   Patients- what is the status of their health?
Fetuses   Native Americans
Non-English speakers (Include copy of all materials in language of participants and certification of the translation and back-translation: http://researchintegrity.asu.edu/humans/forms)

The system is open to the public so anyone could use the system. It is possible that some children under 18 at these public events with their families may desire to use the system.

a) If **any** of the above categories have been checked, please state how you will protect the rights and privacy of these individuals.
I will be observing and taking notes only on user behavior with the system, specifically with regard to how engaged they are. I will not be noting any potentially identifying features such as names or physical appearance on anybody. If I notice someone under 18 is there, no observations will be notated on that person and no audio recordings will be taken.

b) Please provide the rationale for the choice of the subjects including any inclusion criteria.
This is a test of how engaging the interface is, so users self-select and freely choose to participate based on their interest in trying the system. Because the system is installed at arts venues that are interested in showcasing the system to their audiences, the system must be available to all.

c) Will any ethnic/racial or gender groups be excluded from this study? If so, provide the rationale for the exclusion criteria.
No exclusions.

RECRUITMENT

8. Describe the process(es) you will use to **recruit participants** and inform them about their role in the study.

4

a) Will any of the following be used? (Check all that apply and attach copies)
 Internet/Email
 Newspapers/radio/televisi on advertising

 Posters/brochures/letters Cover Letter (attached)
 Other Participants will have access to a musical instrument letting them play along with an electronic music DJ. This system will be installed and performed during several one-day public events in the Mesa area (free museum days, second fridays in Mesa, and so forth). Event-goers will be able to play this game freely if they wish to.

b) Does any member of the research team have a relationship (i.e., teacher, coach, physician, therapist, service provider, etc) with individuals who will be recruited for this study or with institutions that will be used to recruit for this study? If yes, describe this relationship in detail and explain how the research process will avoid any potential problems (e .g, coercion or appearance of possible coercion in recruiting) or conflicts of interest arising from this investigator's dual roles.
No.

DECEPTION

9.   Does the proposed research require that you deceive participants in any way?
No.

a) If your response is "yes," describe the type of **deception** you will use, indicate why it is necessary for this study, and provide a copy of the debriefing script.

COMPENSATION

10. Will any type of compensation be used? (e.g. money, gift, raffle, extra credit, etc)
a)
No (This is a study on engagement. Compensation would distort the results.)

b) Explain why the compensation is reasonable in relation to the experiences of and burden on participants.

c) Is compensation for participation in a study or completion of the study? (Note: participants must be free to quit at any time without penalty including loss of benefits).
 Participation   Completion

d) If any of the participants are economically disadvantaged, describe the manner of compensation and explain why it is fair and not coercive.

INFORMED CONSENT

11. Describe the procedures you will use to **obtain and document informed consent and assent**. **Attach copies of the forms that you will use.** In the case of secondary data, please attach original informed consent or describe below why it has not been included. Fully justify a request for a waiver of written consent or parental consent for minors.
(The ASU IRB website has additional information and sample consent and assent forms.)
We hope this IRB will be considered exempt.

RISKS

12. What are the potential risks of the research? (Check all that apply)
No apparent physical, social, emotional, privacy, criminal, employability, or legal risks.

a) Describe any potential risks to human subjects and the steps that will be taken to reduce the risks. Include any risks to the subject's well-being, privacy, emotions, employability, criminal, and legal status.   Risks are minimal.

Social Behavioral IRB Application Form – Page
Revised April 2011   5

Similar to the risk of playing a Microsoft Kinect video game at home. There is no social stigma against playing video games.

BENEFITS

13a) What are the potential benefits to the <u>individual subject</u>, if any, as a result of being in the study? The system is being installed at events at various public arts facilities with the hope that the audiences there will be engaged by the system; this is why the administrators at these locations were willing to host the system. This is a study of engagement, and we hope participants will find our interface engaging and enjoyable.

b) What are the potential benefits, if any, to others from the study? This study will help us better understand how to design technologies that are engaging to people over a long duration. Such technologies will be helpful in many applications related to achieving long-term goals or overcoming challenges. We intend to eventually use this framework to design long-term engaging interfaces for helping people rehabilitate injuries and disorders.

DATA USE

14. How will the data be used? (Check all that apply)
 Dissertation, Thesis, Publication/journal article, Conferences/presentations

PROTECTION OF CONFIDENTIALITY

15. Describe the steps you will take to ensure the confidentiality of the participants and data.

a) Indicate how you will safeguard data that includes identifying or potentially identifying information (e.g. coding). Identifying information is not necessary for this study. No identifying information will be notated during the system demonstrations. All observations will be focused on the behavior of users and whether they found the system engaging. With regard to audio recordings, no exchange of names is required so identifying features should not be recorded. If a name is accidentally recorded, however, it will be spot-deleted on the first pass through the data.

b) Indicate when identifiers will be separated or removed from the data.
No identifiers used. If any identifiers are accidentally recorded in the audio (such as someone in the background calling a subject by name) these will be erased from the recording upon the first pass-through of the data.

c) Will the study have a master list linking participants' identifying information with study ID codes, and thereby, their data? If so, provide a justification for having a master list. (Note: In many cases, the existence of a master list is the only part of a study that raises it above minimal risk, that is, places participants at risk.)   No.

d)If you have a master list and/or data with identifiers, where on campus will the list and/or data be kept? The game will be run using an ASU computer in an ASU space. **(Data sets with identifiers and master lists, whether electronic or in hard copy, should be securely stored on an ASU campus except in unusual circumstances (e.g., research conducted out of the state or country).)**

e) If you have a master list, when will it be destroyed?
We have no master list. See question 15c.

f) How long do you plan to retain the data?
1 year.

g) How will you dispose of the data?

6

h) Where on campus will you store the signed consent, assent, and parental permission forms
We hope this IRB will be considered exempt.
(If applicable)? (**Consent, assent, and parent permission forms should be securely stored on an ASU campus**)

INVESTIGATOR INTERESTS

16Have all investigator filed a current annual conflict of interest questionnaire with the ASU Office of Research Integrity and Assurance? It is the COEUS module at: http://researchintegrity.asu.edu/coi
Yes

a) Do any of the researchers or their family members, have a financial interest in a business which owns a technology to be studied and/or is sponsoring the research?
No

b) Are there any plans for commercial development related to the findings of this study?
No

c) Will the investigator or a member of the investigator's family financially benefit if the findings are commercialized?
No

d) Will participants financially benefit if the findings are commercialized?
No

BIOLOGICAL MATERIALS

17a) Will biological materials be collected from subjects or given to subjects?
No (If no, please skip to question 18)

TRAINING

18. The research team must document completion of human subjects training from within the past 3   years.
 (**For more information see:** http://researchintegrity.asu.edu/training/humans)

Please provide the date that the PI and co-investigators completed the training and attach the certificate.
Isaac Wallis – 11/01/11
Todd Ingalls - 12/11/11

PRINCIPAL INVESTIGATOR

In making this application, I certify that I have read and understand the ASU Procedures for the Review of Human Subjects Research and that I intend to comply with the letter and spirit of the University Policy. Changes in to the study will be submitted to the IRB for written approval prior to these changes being put into practice. **I also agree and understand that informed consent/assent records of the participants will be kept for at least three (3) years after the completion of the research**. **Attach a copy of the PI's CV unless one is already on file with the Office of Research Integrity and Assurance.**
Name (first, middle initial, last):

Signature: Date:

FOR OFFICE USE:	This application has been reviewed by the Arizona State University IRB: Full Board Review Expedite Categories: Exempt Categories: Approved Deferred Disapproved Project requires review more often than annual Every months
	Signature of IRB Chair/Member: Date:

8

C.3 IRB exemption letter, applicable to original submission and amendment.

ARIZONA STATE UNIVERSITY
Knowledge Enterprise Development

Office of Research Integrity and Assurance

To: Todd Ingalls
MCENT

From: Mark Roosa, Chair
Soc Beh IRB

Date: 12/22/2011

Committee Action: **Exemption Granted**

IRB Action Date: 12/22/2011

IRB Protocol #: 1112007163

Study Title: Long-Tern Engagement with a computer interface

The above-referenced protocol is considered exempt after review by the Institutional Review Board pursuant to Federal regulations, 45 CFR Part 46.101(b)(2) .

This part of the federal regulations requires that the information be recorded by investigators in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects. It is necessary that the information obtained not be such that if disclosed outside the research, it could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.

You should retain a copy of this letter for your records.

660 S Mill Avenue Suite 315 Tempe, AZ 85287-6111 P 480 965 6788 F 480 965 7772

APPENDIX D

AXIAL CODES IN *PDMUSIC* ETHNOGRAPHIC STUDY

D.1   Event Codes

- AUD (Audience Size): Rated from smallest to largest. Space55, the smallest, had approximately 20 potential users. The largest, monOrchid, had at least double the next-largest with approximately 500 potential users.

- DIST (Distractions): AMY event-goers were uniquely distracted because many of them were watching small children. Otherwise, distractions consisted of competing exhibits (as was the case with AMY, IceHouse, and monOrchid) or few nearby distractions (as was the case with Space55 and MCA).

- USB (Usability): *pdMusic* was uniquely beset with crashes and bugs during the Space55 event. These bugs were eliminated in the other events, although usability issues with regard to slow calibration times and occasional short tracking glitches persisted throughout all events.

- TAR (Targeting): Event targeting was rated from worst to best. Space55 was uniquely well-targeted, and the IceHouse was also well-targeted. AMY was uniquely poorly-targeted. This code also encompasses overall novelty, as system novelty varied inversely with event targeting.

- PRM (Promotion): All events were promoted using social media in the form of Facebook events and list server email invitations. Space55 was unique in its promotion via flyers. The other four events were promoted by event managers, but *pdMusic* was not specifically promoted within the events.

- ACC (Accessibility Within Event): All installations happened in the context of some wider event having its own audience. This code describes how near or far *pdMusic* was located to the center-of-activity of that wider event. *pdMusic* was located centrally to three of the events, was within sight of the MCA center-of-activity, and was within walking distance (three city blocks) from First Friday during the Space55 event.

- SIZE (Event Size): This code describes the size of the wider event that *pdMusic* was installed in. Ranges from small (approximately 100 event-goers) to large (thousands of event-goers). The IceHouse event was a special case, as it can be considered an event within an event. *pdMusic* was installed centrally to the SlipStream art exhibition, which was a mid-sized event, but this also occurred within context of First Friday, a large event taking place 1.5 miles away.
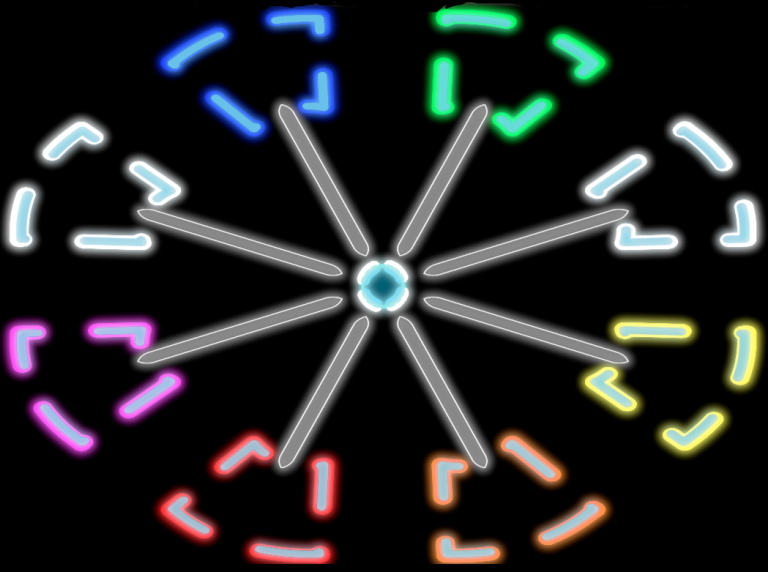
D.2 Subject Codes

- ID (Subject Number): The first user of each event is given index 1, the second, index 2, and so forth.

- DATE (Event Date): The date of each subject's usage. Indicates which event usage took place on.

- G (Gender): The user's gender.

- AGE (Age): The apparent age of the user. Common codings include 'young' (18–40), 'middle-aged' (40–60), and 'older' (60+).

- GRP (Accompanied By): A brief description of people accompanying user (e.g. 'one female companion,' 'three male friends,' 'one child,' and so forth).

- ADT (Reason for Attending): The reason the user came to the event, if known (e.g. 'event staff,' 'event attendee,' 'friend of *pdMusic*,' and so forth).

- P (Time playing): Total (approximate) time spent using the system during this event.

- Obef (Time Observing Before Usage) Approximate time spent observing the system before trying it.

- Oaft (Time Observing After Usage) Total approximate time spent observing the system after trying it the first time (note: if there were multiple usage sessions, this includes time spent observing between sessions).

- NS (Number of Sessions) Number of individual *pdMusic* sessions for user, if multiple sessions were undertaken.

- STP (Reason Stopped) The reason the user stopped using system, if apparent. One common reason was to step aside and let others use system.

- FA (Focused Attention) The extent to which the user seemed absorbed in the activity, and any descriptive indications informing this observation.

- ANX (Performance Anxiety) Any feelings of awkwardness or performance anxiety that were indicated by the user.

- SOC (Sociability) Whether the user exhibited noteworthy friendliness, extroversion, or performative playing styles.

- WUA (Would Use Again?) If, in conversation with a user, the question is asked: "Would you use this system again, if it were installed at some other public event and you were there?", this is the given answer.

- KNO (Knowledge or Interests) Any interests or expertise claimed by the user, especially those which are relevant to the system such as dance, programming, video gaming, and so forth.

- COMP (Compliments) Any compliments the user made on the system. Common ones include calling the system "cool" and similar.

- CRIT (Criticisms) Any criticisms or feature requests made by the user.

- COMM (Comments) Any directed speech that is neither critical nor complimentary.

- VOC (Vocalizations) Non-verbal sounds made during usage such as laughing out loud, screaming, and so forth. Also includes spoken language if it is clearly not directed at any person (e.g. "Wow" or "What the heck?").

- MOV (Movement) Noteworthy movement styles while playing. Includes dancing, rhythmic arm movements over the targets (which was common), and body contortions.

- LEARN (Learning) Any indications, including user comments, that the style of play evolved over time.

- LOCUS (Locus of Perspective) Any comments indicating the user is considering the system from a perspective not his or her own, i.e. "My grandkids would love this!"

- NOTES (Notes) This is for miscellaneous data not encoded elsewhere.

PROMOTIONAL MATERIALS

APPENDIX F

USAGE TIMELINE: SPACE55 EVENT

The first *pdMusic* installation, at Space55, had few users, making it feasible to construct a usage timeline for that event. This appendix lays out the timeline, discussing each user and transcribing some interesting remarks and dialogue. The timeline begins with system installation.

Installation   The system was installed and operational by 7:30 PM, and potential users began appearing at around 8:00 PM, starting with a group of three who came into the space and inquired about the system. Unfortunately, the system crashed before I could demonstrate the system, and they left. Later, another group of four males appeared. One asked, "What kind of games y'all got?" When I launched into my system demonstration, the group quickly made its exit. It is possible they envisioned a different sort of video gaming environment and were uninterested in *pdMusic*. Soon thereafter, subject 1 came to the space.

Subject 1:   Subject 1 was the first system user. A programmer, he was walking by himself among the First Friday crowd when my promoter handed him a flyer and struck up a conversation. After the system was described to him, he decided to go try it out. Unfortunately, he set off in the wrong direction and did not find Space55. So he returned, found the promoter again, and got new directions. When he got to Space55, he observed a system demonstration for a few minutes, then tried a usage session which lasted about ten minutes. He was silently focused on the screen during usage or most of this usage session, but towards the end he made a complaint about the system's interactive response. This complaint turned out to be rooted in the *pdMusic* synthesizer configuration:

 Subject: I can't get the response I want–I just want a better response...

266

Me: You mean the tracking?

Subject: Yeah, it's not responding right.

> *Here the DJ changed the sound being played from a sound with slow attack to a sound with fast attack.*

Subject: Oh, there's the response I wanted! (*stops talking and begins performing radial arm sweeps.*)

I was able to briefly interview subject 1 after his usage session. During this conversation, he asked pointed questions about the programming language I used. When I told him I used Objective-C and Cocoa, he said "Oh, that's kind of like C#, right?" During the interview, he suggested a number of possible improvements to the system based on his knowledge of gaming and software design. His suggestions included: adding scoring to compositions, removing the radial lines, making compositions harder, making sounds louder, and redesigning the system for less arm fatigue. At the end of our interview, I asked if he would use the system again were it installed in hypothetical future public events:

Me: If this system were installed at a public event and you happened to be attending, is this something you would try again?

Subject: Oh yes, it's very interesting. You might want to change the color scheme though.

Me: Oh?

Subject: Yeah, on the flyer the colors look really great, but on the screen...

Me: Not so much?

Subject: Not so much.

Subject 2: Subject 2 became aware of the *pdMusic* installation through the promoter, who gave him a flyer, at the main First Friday thoroughfare. Him, and his female companion, came in during the usage session of subject 1. Subject 2 and his companion sat and watched subject 1 use the system, then took their turn during my interview of subject 1. The female companion took pictures and filmed while subject 2 played. Subject 2 seemed to enjoy the system. He seemed absorbed. He maintained a smile throughout his usage session, which only faded somewhat upon parts of compositions that were technically challenging. He occasionally

groaned in effort as he played. He used rhythmic hand motions when playing notes. These rhythmic motions were related to his only criticism of the system:

 Subject: *[while waving his arm rhythmically over a target]* "I want it to do BPM-tracking *[of the arm rhythm]*. You know, like SMULE."

SMULE is a company that makes mobile apps for electronic music-making[1]. Since it was not widely known by non-musicians, this comment implies previous interest in music technology or electronic music. Other statements and direct claims by this subject also support this observation.

Even though there were system crashes during his usage session, subject 2 played for approximately twenty minutes, after which I gave him a brief interview. During this interview, he asked to see the program code for *pdMusic*, and he also asked for my business card. As with subject 1, I finished by asking if he would use the system again were it installed in hypothetical future public events:

 Me: If this system were installed at an event and you happened to be there, would you be interested in trying again?
 Subject: Yeah, definitely! I don't know why there aren't Kinects *[Kinect applications]* at every DJ show.

Subjects 3 and 4: Subjects 3 and 4 came together, along with one other person (a family member of subject 4) who did not try the system. This group became aware of the installation through promotion via my social network, so they were friends of mine who attended, in part, to see and support my work. Subject 3 had never been exposed to *pdMusic* before, and subject 4 had used it once. Subject 3 used the system for a duration of approximately eight minutes, during which time he asked questions on the operation of the system but otherwise exhibited focused attention. Subject 4 used the system in silence for a period of about ten minutes. I did not formally interview these users during this event.

---

[1] see http://www.smule.com/

Subject 5: Subject 5 became aware of the system through social media. A friend of the DJ, he is a student in the Digital Culture department of Arizona State University. The DJ and I are each affiliated with that department, but I had not previously met this person. Subject 5 was accompanied by one male companion who did not try the system. During usage, subject 5 gave many indications of enjoyment, and seemed to be delighted with every aspect of the system. He made many complimentary statements and would often laugh out loud while playing. He was especially complimentary of the aesthetics and sense of fun built into *pdMusic*. Possibly because of his experience as a DJ, his improvisation style was more tightly coupled to the background music than many other users. For example, he paused during musical breaks, then burst into motion when the music re-entered. He also used rhythmic hand motions while playing.

He played for approximately fifteen minutes. After his usage session, he stayed to observe for approximately fifteen minutes. During this time, he continued to make exclamations and compliments about the system. At one point he exclaimed "That's gorgeous!" in reference to the pink sparkles generated by a correctly-played composition note. As his parting comment, he told the DJ "You're involved in an amazing project here!"

Subject 6: Subject 6 became aware of the *pdMusic* installation through the promoter, who met him at the main First Friday thoroughfare and gave him a flyer. He is an old friend of the *pdMusic* DJ, and recognized her stage name printed on the flyer. He came alone. This subject is a DJ and fire-dancer with an outgoing personality. His style of play within the system was geared toward improvisation, demonstrability, and sociability; a skilled dancer, he seemed to love showing off to the audience. He did not seem to rely on the interface. Once, upon a system crash, he kept dancing as though nothing had happened. After becoming familiar with the interface, subject 6 was able to improvise without looking; he began dancing around in circles and facing away from the screen while playing. He clearly preferred improvising to

playing compositions, and soon began to ignore the note-lines of compositions entirely, instead playing whatever he wanted to.

Subject 6 had two usage sessions. The first was roughly twenty minutes long, followed by a break of around ten minutes while I interviewed him, followed by the second usage session which was eight minutes long. He was generally complimentary of the system; however, he felt the motion tracking could be improved, and for this reason he preferred slow-attack synth sounds as they require less exact rhythmic timing. He remarked to another user: "The timing is the hardest thing about this system. But once you've got that, it's great!" During his interview, I asked if he would use the system again if it were installed in a hypothetical future event:

Me: If you had a chance to use this system again at some other event in the future, is it something you think you would do?

Subject: Oh, yeah! Hey! If you want, I can get some people from the dance community together to sort of focus-group it!

Subject 7: Subject 7 became aware of the installation through a flyer, given to him by the *pdMusic* promoter, on the main First Friday thoroughfare. He came with a female companion. Even though he came specifically to see *pdMusic*, he was initially reluctant to try it himself. He expressed discomfort at using the system in front of people because he did not know how to use it. His companion urged him to try, telling him "This looks like something you would really like." Finally, after a demonstration on how to use the system, subject 7 decided to try. He played in silence for approximately five minutes, then abruptly ended his session by declaring "I kind of feel weird playing this. It is super-fun, though." Subject 7 and his companion left immediately thereafter, before an interview could take place.

APPENDIX G

TRANSCRIPTS OF SELECTED INTERVIEWS

G.1 DJ Interview

Me: So, as you were saying before I started recording, there's a problem with a suspension, with—everything freezes when either of us changed something on the interface.

DJ: Right. Well, the key seems to be fine. It's just when triggering a game that it happens. Right?

Me: Oh Really? If so, then that'll help me debug that. Changing keys doesn't freeze it?

DJ: I don't think it does.

Me: I will definitely check that... 'cuz if that's the case then it might be something very simple that's causing the problem.

DJ: I mean I'm not 100% sure but the last time we used it I thought that was how it was working.

Me: (*moving notes around on the table*) …alright …are there any other thoughts you've had on *pdMusic*?

DJ: Well, I think that people liked it and they tended to interact with it—one of the things, especially from the last performance, was that I felt like the game aspect of it was a plus and also a minus, because there were some people that weren't interested in playing the game. Like there was one older woman that was just free-forming, you know playing along with the music, and then as soon as I trigger the game—I didn't ask her, I just triggered it—then she walked off. And I asked her, I was like "Oh you don't want to play a game?" And she was like "No." (*laughs*)

Me: Ooohhh. (*laughs along*) So, do you think most people liked improvising more than games?

DJ: (*long pause*) That's my inclination I don't... I don't know if I've been paying enough attention to everyone to say for sure one way or the other, but it seemed like the game was one aspect, but the people that played the longest and got into it the most, seemed to really enjoy just improvising.

Me: Ok. That's interesting. (*long pause*) Can you think of a reason why that might be?

DJ: Well, I think in that setting, the idea of, like, the pressure to perform in that environment in front of so many people... that might be an inhibitor for some people.

Me: Yeah, that makes sense. So the games add more pressure to perform. Do you think they'd care as much if they didn't have people watching?

DJ: See, I would tend to think not. Like, if it wasn't in such a high-traffic area and there weren't so many people around, I feel like that would be less of an issue.

Me: Ok. Yeah that's how I would sort of–

271

DJ: –Yeah, but at the same time, it varies, because if there were people that were in groups and sort of had somebody instigating them to, like, play with the system, then I feel like the game mode wasn't so much of an issue. But if it was somebody that just happened to be walking by, and was curious and wanted to try it out, then they might just try it and not use it as long.

Me: Yeah, so when people go in groups like that—when you see people coming up, and they're accompanied by other people—and I'll ask you specifically about one case I remember here in a minute—did you notice any differences in the way they would behave?

DJ: If they went as a group versus individual?

Me: Yeah. If somebody came up and they were alone, versus if somebody came up and they were with their partner, versus if somebody came up and they had several buddies with them.

DJ: Yeah, I think there was a sort of spectrum that happened. If they were a couple for instance, then one might encourage the other to go try the system. Or, a lot of times people wanted to play the game together. So that might be an interesting thing to consider for a future edition.

Me: To do a two-player version?

DJ: Yeah, a two player version.

Me: I really want to. I think that is one of the things that really needs to happen if I was going to make a super-version of this. Which I probably will end up doing, eventually. It might take me a year or something. (*long pause*) And, so, if they had friends along?

DJ: If they had friends along it was more, uh... I felt there were times it could be competitive, like they would encourage each other to go out *[to use the system]* but then be like "Oh I'm gonna try to do better than you!" (*laughs*)

Me: Ok. (*laughs along*) Yeah. Did you ever see instances where it was like, because they came in groups, they really didn't get into it?

DJ: I don't know if they didn't really get into it, but there were times when people came in groups but then fulfill more of a spectator role as opposed to interacting with it, because somebody else was using it.

Me: Yeah, I would agree with that. So, at the third show we did—that was Mesa Center of the Arts—your husband *[Subject 12]* and some of your friends showed up towards the end of the night. And your one female friend, *[Subject 13]*, she didn't really want to play, and I think her husband *[Subject 11]* was like "You should try this." And then she played it, but she just played it for maybe a minute and a half and then stopped.

DJ: I think sometimes people come—I don't know, I think she wanted to come more to support and hang out than to engage with the system. But she succumbed to the peer pressure and then was like "Ok, I'll try."

Me: That makes total sense. I just assumed that she might be someone who viewed a system like that as just inciting performance anxiety more than anything else.

DJ: It's possible. I didn't really have the chance to ask her. But I know her, and we've had Rock Band sessions[1] so... I mean of course that's in a smaller group of people, but... for her I don't think that was necessarily the case.

Me: Ok, so she was just there, like you were saying. Now, in keeping with this line of questioning I'm on... The first night I noticed—probably because we both put out Facebook announcements—that you had a couple friends who showed up that were from the DJ community. There was *[Subject 5]*, and then there was the fire-dancer guy. And those guys both came because they saw your Facebook post, right?

DJ: Well, I think *[Subject 6]*, the fire-dancer, I think he ran into Kristi *[who promoted that event at the First Friday district]*.

Me: Oh! and he recognized your name from the flyer?

DJ: Yeah, we had done shows together years ago.

Me: Ok, that's cool. And... so that night was so crazy, it was our first show. And there were all those bugs and things happening. It was very hard to do a good job of data collection on my part. Although, I did get some notes, especially on people that used the system a long time, and both of those guys did use it a pretty long time. But I'm just curious if you can relate anything they might have said.

DJ: Well, again, going back to some of the technical aspects... because *[Subject 5]* has some experience with design work and computer music interaction, he was commenting on the smoothness of the tracking and the graphics.

Me: He was thinking they weren't smooth?

DJ: No, that they were. He was impressed that, visually, it tracked really well without any type of artifacts.

Me: Yeah, ok. That's good. (*brief chuckle*) Yeah, I remember he seemed to really like the aesthetics. And I remember *[Subject 6]* was sort of talking about dance, looking at it from the perspective of dance.

DJ: Right, yeah. I think, what he said and what I saw with a couple people that really interacted for a lengthy amount of time in the system was that they would incorporate, like, dance choreography into the playing of the game. So that was kind of fun for them but also I felt like maybe it was a little bit limited, like they would have liked it if there was more integration between dance movement and—in terms of the gestural movement to control the sound—if that could be more integrated with dance movements.

Me: Ok. So dance choreography better integrated into system, would help those guys.

DJ: Yeah.

Me: That makes sense. And a lot of people—a lot of people as I recall—wanted some kind of... some people didn't even know they wanted it, they just thought it existed, but some kind of rhythmic thing over the targets? They really wanted that rhythmic movement to do something, it seemed like. In every event, there was somebody who tried that. Doing this kind of stuff (*Here I moved my arm rhythmically in place as if I were touching a note target in the system*)

---

[1] see www.rockband.com

DJ: Which, there was one particular guy at the last show who did that a lot, and it worked well for the piano and the shorter sounds, but not for the longer sustained sounds.

Me: Oh, yeah, yeah, yeah.

DJ: So I kind of kept going back to either the piano or one of the other shorter keyboard sounds to facilitate that.

Me: So, you really liked the piano sounds, just in general. Most of the time, you went for the piano, I noticed.

DJ: Because I felt it was the most responsive.

Me: Ok.

DJ: And people could make the connection and kind of get a foot into the game. (*laughing*) That's the gateway sound, I guess.

Me: That's a—gateway sound, that's a good turn of phrase. I'm going to pause this and write that down.

> *At this point the interview was paused for a short period to manage note-taking and render the recorded audio. It resumed momentarily.*

DJ: And continuing on with the piano sound? Compared to all of the other sounds in the library, that one was the loudest.

Me: Ohhh.

DJ: Yeah, that one was definitely the loudest, but also seemed to be audible in most songs, like I didn't have a problem running into the same frequencies, where with some of the other sounds I had to EQ it and increase the volume so I guess one thing for future iterations is a way to normalize the presets in Crystal Synth.

Me: A way to normalize the presets... yeah. That would be good. The best way to do it would be to actually go through and—I'd need a few more people, a team to help me to do it—but we'd go through and get the best sounds that we like the most and then get them all perfect. Kind of master them *[e.g. like mastering an album]*, do a mastering of the different sounds, but within Crystal. (*long pause*) So... other sounds, were there other sounds that you found–

DJ: –Well, I used the bass guitar sound that I made, it was more of a jazz bass guitar, so that seemed to work, but again, the first time I used it, making the transition from other sounds to that sound, there was a large volume difference.

Me: What about the slow sounds?

DJ: I think the slow sounds worked well for specific songs, and also for people that weren't trying to push the interactivity of the system, if that makes sense. Like we were talking about how people were trying to do rhythmic things with it? That obviously didn't work as well for the longer sustained sounds.

Me: If they were made to be louder and compete with the piano sound in all those ways, would the slow sounds be better as gateway sounds, do you think? Or is piano still the best thing?

DJ: I think piano is the best because of the responsiveness.

Me: Ok.

DJ: So I feel like the longer sustained sounds are kind of more fun for music composition as opposed to game interaction.

Me: Ok, yeah, that makes sense to me. So they might have been really good for improv stuff. I saw a few people–

DJ: –Like, especially, some of the chordal sweeps that would happen? If they did that it would sound really good, and people would respond to that.

Me: Ok. Now, I know you've got a time limit, so I want to move into what your experience was as DJ using the system.

DJ: Ok. For me, I had to shift the way that I DJ for this system. In a couple of different ways. Partly because of the separation of cognitive load, because I had to be focusing on what they were doing and also facilitating the interaction with the system, and I'm doing the DJing. Now, normally when I'm DJing and I'm just focusing on my own setup I tend to do more EQ work, cutting—more like, I guess, DJ tricks (*brief chuckle*). But partially because of when the user decided to interact with the system based on the point of the song that the track was at, I couldn't always transition in a way that I normally would. Which resulted in, maybe, some of the songs were played longer than if I was just a DJ.

Me: Ooooohhh. Ok. That's really interesting. So doing this system made you play songs longer, made you play... more simply, I guess?

DJ: So the focus was less on the technique and artistry of DJing and more of an accompaniment to the system and the interaction.

Me: Did you find that challenging?

DJ: (*laughing*) At first, yeah.

Me: (*laughs along*)

DJ: I mean it's definitely different to have your attention so divided.

Me: Yeah. If it were somehow to—like, is it partially divided because there's two different computers you are having to run?

DJ: It's partially that, but I think if there was a way that we could have automated the key detection, then that would have made it a lot easier.

Me: Ok. Key detection.

DJ: And here's why. Part of the time, especially in transitioning between songs of different keys, there's a point—especially if the user is interacting during that time—there's a point where I have to make a decision of when am I gonna switch the key on them, and how is that gonna affect the system. Because I don't remember if there was a glitch there or not. I would always have to negotiate that transition with factoring in all these different elements.

Me: Yeah, that does make sense, and I think a future version would need automatic key detection. With the reduction in your ability to perform DJ tricks, did you think it got too simple? What I mean by too simple is that, at some point a DJ isn't even necessary, you could just have a playlist. Was it ever to that level of simplicity?

DJ: I don't think it was ever to that extent. There wasn't as much interaction, on that level, that I would have liked. To be able to kind of improvise with the player?

Me: Mm-hm.

DJ: It would have been nice if that could have happened more.

Me: I understand that.

DJ: So I mean there was still taking into consideration of "Ok they chose this song to start playing on, and where am I going to take this and hopefully keep them."

Me: Yes.

DJ: So I mean there was always that, and I don't know if a playlist could ever substitute for that in that regard. And I would still try to do a couple things but it really just depended on the timing of the interaction based on trying to do all these things.

Me: That makes total sense. So you've had a lot of experience in DJing. Do you see systems like this in DJ setups, clubs and so forth? Does this sort of interaction seem like something you could see infiltrating the club in the future?

DJ: Yeah, I definitely see the potential for that application. In my mind, there would have to be a few tweaks to it, like most of the things that we've already talked about. The first being more integration with the way DJs do things now, so that there's less attention required that detracts from what the DJ normally does. And then, for the user perspective like in the club dance environment, making the gestural interaction more dance-related would definitely help. I think if that was developed a little bit further to also include not just the hands, but also the feet, which I saw people trying to do anyway (*laughing*). I feel like that could be a really good system that could be integrated with go-go dancers, for instance.

Me: So one way that this could be integrated with some types of DJ equipment would be that maybe a plug-in could be made, like almost a VST-style plug-in could be made that interfaces with the Kinect and does all the stuff that this does. Of course, handling all the video projection kind of exceeds the bounds of what a plug-in normally does, but is that the kind of integration you mean? Have it built right into people's interfaces?

DJ: No, I just meant integrating it into the performance-slash-spectacle of the club environment. I think it may be possible to have that type of integration, but for now I feel it is better off as a separate app. The main thing would be to somehow automate the key detection and then make the sounds controlled by the user instead of the DJ. But maybe have a way that the DJ could pre-select which sounds he wants to be available.

Me: Yeah.

DJ: (*laughing*) If that makes sense.

Me: Yeah! It totally makes sense. (*long pause*) Ok, so I want to go back and talk again about ways that you needed to DJ differently, and ways you needed to think about this differently. So what we talked about before was that you needed to play songs longer and negotiate the transitions–

DJ: –Well, particularly if they were playing a game, I couldn't transition songs while they were playing games.

Me: Ohhhh

DJ: Because that would just be weird. Like, in my mind, you're playing along to a song, and then you're playing along to a new song. Which would be fine if you were just improvising, but because you're playing a game, I felt like that would just be too weird. So maybe it would have been ok, but just from my perspective, I tried not to do that to people.

Me: Ok. Yeah, that makes sense. Now, those were sort of challenges or I guess you could call them less desirable things about interacting with this system as opposed to just being a regular DJ. Was there any aspect of this which was sort of skill-building for you?

DJ: Um... (*chuckles wryly*) multitasking. That's one. (*pause*) I mean, part of it was observing different people and trying to still find a way to engage with people even while trying to manage other things. Finding it really fulfilling when people would stay and play a long time, especially if they would play a few games but also were improvising. And being able to do a little more of that kind of exchange, was definitely fun.

Me: Ok. There were a few users that I saw that... they seemed especially sociable, I guess you might say, or maybe they were kind of like, showing off.

DJ: Performers, yeah. They had the performer personality, yes.

Me: (*laughing*) Performer personality. I'm just gonna write that down because that's another really good turn of phrase. Was there anything you noticed about that?

DJ: Well, I think there were two groups of that. There were some that were more just in it for the attention, whereas I felt that some people had a certain knack for the musicality of it, and so their playing was more musically influenced as opposed to "Hey, look at me!" (*laughing*)

Me: (*laughs along*) Oh, hey! Can you expound for a little bit on some differences between the different events?

DJ: Ok. Um. For sure, the first event at Space55... just in terms of the amount of traffic, but also the stage construct was not conducive to what we were trying to do.

Me: Ok. And it wasn't conducive because–

DJ: –Well, it created—it sets up this performer/spectator relationship. And when we're trying to get people to interact with the system that's not what we want.

Me: Yeah, I can buy that. And we obviously didn't get that many people because we were so far off the beaten path for the rest of first Friday. In a way, that was one of my favorite shows, because the people we did get were mostly really into it.

DJ: Yeah.

Me: So, that was the first show. Stage not conducive. So, for the second show, which was the next morning, if you recall.

DJ: Yeah. That was at the children's art museum.

Me: What were your thoughts on that show?

DJ: Well, definitely we had a different audience. Based on that setting, we were kind of competing against all the other exhibits for people to interact with our system so... yeah, I felt like the level of engagement wasn't as good as the first one was.

Me: Level of engagement not as good. And you just chalk that up to different audience?

DJ: Well, different audience, and also the venue. And the reason for people going, like it was primarily families who were taking their kids for an outing, so the attention span of children is much shorter than adults so...

Me: Yeah. I definitely agree with that. So... Mesa Arts Center.

DJ: Uh... A, I think we had the weather going against us.

Me: Oh, yeah, we had a little rain that night, didn't we? Yeah, so in your opinion that had the effect of...

DJ: I think there were less people that came out to the event than we would normally have.

Me: Yeah. (*pause*) I kind of view the two Mesa shows—I kind of group them. Do you do the same?

DJ: I do because I feel like we were kind of part of something else but people weren't necessarily coming to see us, and it still felt like there were a lot of families, and we were kind of on the way to something else.

Me: Uh, right. They weren't coming to see us, they were going to something else, and we were just there too, right?

DJ: Right. And that's different than the first and the last show.

Me: Ok. And the last event? I know we've talked about that one a little already, but in the context of these other shows...

DJ: I felt like, in general, the audience was more interested in participating.

Me: That's how I felt too, yeah.

DJ: We did a little experiment at the end when the drum circle got kinda loud and just let people improvise over the top of that. And I felt like people played for a shorter amount of time than they did when they played with the DJ.

Me: There was one guy *[Subject 26]* who jammed out on the Debussey sound for quite some time...

DJ: (*laughing*) Well, I think he jammed out on many sounds.

Me: Oh, you're thinking of—there were two guys–

DJ: I'm thinking of the one guy who was totally into it *[Subject 9]*. Uh, I gave him the piano sound a few times and then–

Me: Yeah, he was, like pretty funny? Joking with everybody?

DJ: Yeah.

Me: Yeah. Yeah, he was probably my favorite user of all four events.

DJ: (*laughing*) Me too.

Me: But yeah, later on, after he was done—you had stopped but the system was still running with the drum circle—and some other guy got up there and we had the debussey sound and he just improvised on that for five or eight minutes. And then walked off. We were not actually even hanging out by the system at that time. Or at least I wasn't.

DJ: Did he have a musical background? Did you ask him?

278

Me: Did I ask him?

DJ: I'm just curious. Because I felt like there were a couple of other users that had some...

Me: I didn't really find that out from enough people. Maybe in the next show we'll get some really dedicated users like that and we'll–

DJ: –Because I thought that there was one guy who seemed to get it–like early on–where he was asking you the sequence of the notes and was curious about how this was musically structured.

Me: Early on in that show?

DJ: Yeah, in the last show.

Me: There was. There was an older gentleman... *[Subject 5]*

DJ: Yeah, I think that's who I'm thinking of.

Me: Yeah he's definitely got... he knew some musical terminology that you wouldn't get outside of having some training. So yeah, I would say he's got some musical background. That guy did, but I really didn't even get to talk to the guy who did the improvisation at the end of the night. And, I did talk to the really into-it one, but I wasn't trying to lead the conversation too much, and he never volunteered that he was a musician or anything like that.

DJ: I think maybe we should add that onto our questionnaire.

Me: Yeah, I will do that.

DJ: And maybe for people that are interested, you know how you kind of wrote up the sequence of notes for me? Maybe have that for people that express that level of interest.

Me: Yeah, I should probably put something like that right up on the interface that indicates what scale degree each target is...

DJ: I feel like that would enhance the improvisation. Because people that have some musical knowledge would be like, "Ok I can make a melody, I can produce something that I want to" as opposed to not being really sure what is happening, like "This sounds good but I don't know what it is."

Me: One thing different about that show was that I turned on the silhouette hoping to draw in more people. Did you have any thoughts on that?

DJ: Well, it just adds another level of visual feedback. It allows people to say "Oh, that's me."

> *At this point, the DJ needed to attend another meeting, so the interview was called to a close.*

G.2   An Expert User

Me: I know you attended, because—well, you might have been coming to First Friday anyway, but...

Subject: Because I know someone who knows you.

Me: Right. You probably played for, I don't know, five minutes, right?

Subject: Yeah, seems about right.

Me: Ok. So, do you have a background in music?

Subject: No. I played English handbells in high school for about three years but I had to have the notes written on the music.

Me: Alright. And, you're not a dancer?

Subject: No. I'm a climber so I probably do have good proprioception.

Me: And I know you do kind of engineering and media design and things like that.

Subject: I'm a software engineer and I'm very interested in human-computer interfaces

Me: Gotcha. Alright. So, do you have any comments about the system? Anything you noticed about it?

Subject: Ok, a couple. First observation, now that I see someone else playing it, is that, if they're more musical then they're getting a different and more musical result. They're driving toward something different than what I knew to drive for. When I first experienced it, I felt like I needed to go slow enough for the system to keep up with me? And it wasn't actually true. Later on, I figured out that I could swirl my arms around and it could keep up. So I don't know what about the system made me think I had to go slow, but I did think that for a long time.

Me: Ok, cool. Did you find yourself getting into it?

Subject: Uhhm. No. (*pause*) Maybe when I did the swirly thing. When I was going faster I was into it more. And I wanted to, kind of move my whole body, rather than just my arms. And actually I felt like if I could move my body and it would kind of follow me and maybe it had some centering thing, and the centering thing kept my arms in the same radius.

Me: Ok, another quick question. I'm just curious about—well this is a performative system, and that could be a good thing for some people and maybe a bad thing for other people, so I'm curious how you feel about that. So some people would find a system like this to be inducing performance anxiety more than anything else, and of course other people really like doing it in front of other people.

Subject: So, what do you mean by performative? Expand on that?

Me: Well, I just mean that you're generating music in real-time, you're kind of improvising it, and right now, there's not too many people standing around but sometimes there might be–

Subject: –Often there is. So for me, then, I didn't think of it as performance at all—or hardly at all—because there didn't happen to be an audience. So I didn't have any of those thoughts coming up. And probably the length of time was probably geared more

by "I wanna see if something else is gonna happen." So only two things happened, I went through several iterations I could either do free-form improvising, or you had one game. So once there was one game, I wondered if there were more games.

Me: So you wanted more games?

Subject: I just wondered, I wondered if there was something else gonna happen. And, I just want to say that the text was way too fast. And when I'm not performing I'll look at it again—I have to wear glasses for far away—so I'll have to see if I could read it well enough, considering.

Me: Gotcha. If the sound had changed more? If the length of time you were in it had to with the extent to which you felt there was more stuff to explore, so if the sound had changed more.

Subject: Let's see. What I think you're asking is "What would have made me stay longer?" And, yeah, a different game, or some sort of feedback on the game about whether I did it right. It kind of is there, like there's the music; but maybe if it highlighted green or something—I don't love that that is my tendency, but it is. Something else that would have been interesting is if I could have picked my own sounds, or done something to actively change the sounds or actively made something different.

Me: Did you find that, by the end of it—you already said that you learned, through doing it that you could move more quickly, so that question is already answered, but—

Subject: Did I learn anything else? From the beginning of my experience to the end, what did I discover that might have changed the whole experience? Umm. I'm learning more from watching someone else do it. I learned the distance, how far I needed to be within the view. I learned how far it would follow me. I learned whether or not it really mattered if I got off-center—and it didn't really matter, which was reassuring, that I didn't have to keep all three things, my arms and my center, in the right place. It was useful that I could drift. It seemed to still work if I moved forward or back. There wasn't just one little teeny square that I had to be in. That was useful. Umm, nothing else I learned comes to mind.

Me: Ok, this question might be a pretty quick one. If there were another event, and this thing happened to be there. Having tried it this time, is it something that you'd be likely to use again?

Subject: Hmm. That's a good question. (*pause*) No, I don't play many video games and I don't do much like this. I can see my husband messing with it. Maybe if there was a recognizable melody, and I could see that my notes were matching the melody?

281

BIOGRAPHICAL SKETCH

For Isaac Wallis, this document is the culmination of a Ph.D. in Media Arts and Sciences at Arizona State University. He has a B.Mus. in Music Synthesis from Berklee College of Music and an M.Mus. in Music Composition, with a concentration in Interdisciplinary Digital Media and Performance, from Arizona State University. His research interests lie in exploring the field of serious leisure and hobby in order to design interfaces that are more engaging and valuable to use. In his own hobbies, Isaac plays harmonica and ukulele, does yo-yo tricks, collects comic books, hacks software and hardware, and runs long distances. Isaac lives with two cats, a dog, a pair of turtles, and his wife, Kristi.