Communicating Meaning in Context-Aware System Design

Jared S. Bauer

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Julie Kientz, Chair

Sean Munson

Wanda Pratt

Program Authorized to Offer Degree:

Information Science

University of Washington

Abstract

Communicating Meaning in Context-Aware System Design

Jared S. Bauer

Chair of the Supervisory Committee:
Associate Professor Julie Kientz
Human Centered Design and Engineering

Computers play an increasingly personal role in our daily lives. Rather than interacting with computers primarily at work or for specialized tasks, computers are now present in virtually all our daily activities. The increased prevalence of technology in our lives has encouraged a rethinking of how technology is designed and developed. Where the focus of human-computer interaction was once on improving efficiency, there is growing awareness that the focus should shift to the user's relationship with technology. This shift has also emphasized the importance of understanding the situated nature of meaning and meaning making. This reorientation toward our relationship with technology and meaning-making leaves open an important question, "how can the user's meaning be conveyed to the technologies we design?"

In my dissertation, I explored this question in three ways. First, I empirically study the practices of context-aware system designers. Second, I conducted a lab study investigating how individuals create context-specific music recommendations. Finally, building off the findings from the lab study, I designed, developed, and evaluated a novel interface type that helps users personalize context-aware music recommendations. Findings from this dissertation illustrate strategies that users draw on in meaning-making, as well as the mismatches between these strategies and the metadata systems rely on to respond. Furthermore, this dissertation contributes insight into the design of systems that enable users to negotiate the relationship between their meaning and context.

ACKNOWLEDGEMENTS

A dissertation is never completed alone, and I am immensely grateful to all who supported me in my work. In particular, I would like to thank my committee members Wanda Pratt and Sean Munson for their support and guidance mentoring me throughout my PhD studies. Additionally, I would like to thank Mark Newman and Ed Hutchins whose enthusiasm for research encouraged me to become a PhD student. Finally, I would like thank Julie Kientz for her support throughout my time as a graduate student.

My research was also greatly improved by my collaborators Jesse Cirimele, Alex Jansen and Aubury Jellenek. Working with Jess and Alex was one of the highlights of my time as PhD students. While it was not included in my dissertation, our work was very influential on much of my research moving forward and I am privileged to have had the opportunity to work with them. I was also incredibly fortunate to have the chance to work with Aubury. She was a consistent source of support and I cannot thank her enough for her dedication.

In addition to my committee members and collaborators, my research has been greatly shaped by my time at Intel’s Interaction and Experience Research Lab (IXR). My mentor, Glen Anderson, and IXR researchers Lisa Kleinman and Jose Sia taught me not only about the mechanics of conducting principled design research, but also how to identify new areas of research and how to effectively communicate my work. I feel that I grew more as a designer during my time at IXR than at any other point in my academic career and I am deeply thankful to them for their help.

I would like to thank my friends, colleges and lab mates Amanda Menking, Lassana Magassa, John Porter, Mia Suh, Alexis Hiniker, Hyewon Suh, Elena Agapie, Ray Hong, Dawn Sakaguchi-Tang, Arpita Bhattacharya and Kiley Sobel. I am truly lucky to have been surrounded by such an excellent community of academics. They have been invaluable as friends and as sources of intellectual inspirations.

Finally, I want to thank my wife Anna Bauer for her consistent support and patience during my time as a PhD student. I have never known a kinder person and I don’t know how I would have done it without her.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. INTRODUCTION

Computers play an increasingly personal role in the daily lives of many. Rather than interacting with computers primarily at work or for specialized tasks, computers are now present in virtually every activity many engage in throughout the day. The increased prevalence of technology in our lives has encouraged a rethinking of how technology is designed and developed. Where the focus of HCI was once on improving efficiency while interacting with technology, there is growing awareness that the focus should shift to the user's relationship with technology (Sellen, Rogers, Harper, & Rodden, 2009). This change of focus is referred to as the third paradigm of HCI or Third Wave HCI; it emphasizes the situated nature of meaning and meaning making (Harrison, Tatar, & Sengers, 2007). This reorientation toward our relationship with technology by emphasizing situatedness and meaning making is an important step, but it leaves open the question; "how can the user's meaning be conveyed to the technologies we design?"

In my dissertation, I explore this question to determine how context-aware system can better enable users to express meaning to context-aware systems. My argument is that current design practices frame the user's context on the designer's terms, which may not reflect how the user expect the system to respond. By enabling users to communicate the relationship between their context and what their context means to them, we can improve the quality of context-specific recommendations. To explore this topic, I have conducted research on the practices of context-aware system designers (Chapter 3), explored how users choose context-specific music recommendations through a chat-client music recommender that I designed, developed and evaluated (Chapter 4), conducted an online evaluation of interfaces for providing context-appropriate music recommendations (Chapter 5), and finally designed, developed and evaluated a context-aware music recommender that allows users to communicate what music recommendations are appropriate for their current context (Chapter 6).

Before discussing my research and its contribution to the field of HCI, I will briefly review the literature that pertains to meaning making and how this relates to current trends in HCI. In doing so, I will motivate why it is increasingly important to enable users to improve how technology responds to user's values, perspective and context.

1.1 THIRD WAVE HCI

The goal of Third Wave HCI is to support situated action in the world (Harrison et al., 2007). The justification for this reorientation is that as we move away from using computers for specific discrete tasks, we expect computers to support more continuous non-goal oriented uses (Leung, 2015). For example, when we browse Facebook, we may not have a specific task we hope to complete (Raacke & Bonds-Raacke, 2008). Instead, we hope to remain connected with members of our social circle, or simply be entertained. To achieve the goal of supporting situated action in the world, and thereby improve our relationship with technology, requires understanding users on their own terms. This understanding has two components: the users' values and perspective, as well as the meaning of their context—their environment, social situation and interaction with the environment. Through an improved understanding of these two components, we can improve user's ability to communicate their meaning to technology.

Allowing the user to communicate their perspective on the meaning of context has been explored, most notably, in the field of Affective Computing. Boehner and her collaborators have investigated how affect can be viewed as *interaction* rather than *information* (Boehner, DePaula, Dourish, & Sengers, 2005). Boehner et al argue that most affective computing systems treat emotion as information, which work by detecting aspects of the user's physiology or environment. This reduces the actual complexity of emotion to make emotion easily transferred from human to computer, or vice versa. Instead, Boehner et al. suggest that emotions are culturally grounded, dynamically experienced, and to some degree, constructed in action and interaction. They argue that by adopting an interactional perspective, systems will better support users in understanding, interpreting, and experiencing emotion in its full complexity. The interactional view of affect encourages researchers to reorient questions in affective computing toward how computers enable users to express their emotion, rather than how accurately they can detect emotion.

Boehner and her collaborators' work demonstrates the value of leaving context open to interpretation by the user, but the other side of the equation is how meaning is imposed on context. For example, the application MoodScope (Likamwa, Yunxin, Lane, & Zhong, 2013) analyzes the communication history and application usage patterns to "measure the mental state of the user" (Likamwa et al., 2013). Likamwa and his collaborators claim that they can detect the user's affective state with 66% accuracy. They go on to suggest that the user's mood could be used "to enhance recommendation systems employed by services such as Netflix."[1] The work on MoodScope clearly views affect as information, but the inability to negotiate meaning with the system makes the information harder to use. If we assume that the application can accurately detect the user's emotional state, what does that mean for the movies it would recommend? Should the application attempt to cheer them up by recommending happier films when they are in a bad mood, or should it choose something more somber that matches their mood? How should the mood of a film correspond to the behavior of the user at all? In fact, what does it mean for a film to be "happy" in the first place? What constitutes the correct response to happiness relies on culturally and contextually dependent notions of "appropriateness". As technology detects and responds to increasingly personal information in users' lives, it is necessary that the system can learn to see the world in a way that corresponds to the user's perspective.

Communicating perspective is complicated by the fact that knowledge is socially situated and computers do not fundamentally understand culture. A computer's "perspective" on cultural is entirely constrained by the designers and engineers that develop the computer. When we interact with technology, there is no preference setting to accommodate the user's socially situated perspective. While work on personalizing recommendations is being done on a variety of fronts, there is little work exploring how to design interfaces that enable users to negotiate meaning and thereby improve how technology relates to the user. To illustrate this challenge, imagine if you had a friend who could not understand your perspective or your culture, and instead used common algorithms present in recommendation systems to navigate social interactions. If you asked them to suggest something fun to do, they would make recommendations based on things you had already done, or things that other people thought were fun. While this might work, it would leave you feeling that your friend did not understand you at all. If you protested and said, "No, I'm not in the mood to do what we did yesterday, I want to do something more exciting" they would be unable to make a better suggestion because they do not know what the culturally relevant meaning of "exciting". More importantly, they do not know what things *you* think are fun and exciting. As language processing and recommendation techniques evolve, the recommendations would undoubtedly improve, but the primary difficulty comes from their inability to understand what fun

[1] http://www.techworld.com.au/article/466164/next_frontier_application_context-awareness_mood/

is and how that relates to the user and to the context. As technology plays an increasingly prevalent role in our lives, it is incumbent on designers to develop new, effective ways to help users communicate their perspective to technology. To provide some background on how knowledge is communicated between individuals, in the following section I will provide a brief overview and functional definition for *constructs*-- the commonsense knowledge and practical reasoning that individuals employ while interacting, and they are essential to achieving mutual understanding (Garfinkel, 1967). I then will connect this notion to work in the field of HCI on context-aware computing.

1.2 CONSTRUCTS AND CONTEXT

Despite how widely constructs are used in the social sciences there is little agreement on how they should be defined or applied. One inherent challenge in defining constructs is arriving at a set of criteria that when satisfied allows researchers to say that a given construct is adequately described. Gibbs details these challenges in his work on sociological theory construction (Gibbs, 1972). Gibbs argues that constructs are inherently incomplete because, even in the theorist's own thinking he or she cannot propose a definition that "closes all meanings of a term" (Gibbs, 1972, p. 124). He goes on to argue that even if a formula could be created to satisfactorily define a given construct, gathering the request data to inform the formula would likely be impossible. To illustrate this point, he discusses the dilemma that might arise when trying to compare international differences in social stratification. Gibbs goes on to argue that the theorist essentially choose between one of two undesirable outcomes: they can either use a term to designate an undefined notion, or rely on a gross oversimplification of the construct.

Given this presumed incompleteness, a definition for constructs may seem hard to agree upon. Despite this challenge, constructs are widely used in conversation analysis. Conversation Analysis (Goodwin & Heritage, 1990), has focused on how meaning is created and communicated between individuals. Conversation analysis is an interdisciplinary field, which has its roots in ethnomethodology and phenomenology. It emerged in the 1960s as a response to primarily cognitive accounts of behavior—such as Gibbs'—which were dominant in the social sciences at the time. In Conversation Analysis, the idea of constructs builds upon the work of Austrian, Philosopher, Psychologist and phenomenological writer Alfred Schutz. Schutz argues that, constructs are the commonsense knowledge and practical reasoning that individuals employ while interacting and that they help individuals "find their bearings with their natural and socio-cultural environment and to come to terms with it" (Schutz, 1962). Garfinkel contributed to this notion by arguing that constructs are vital to achieving mutual understanding (Garfinkel, 1967). In contrast to cognitivist accounts of behavior, conversation analysis placed an emphasis on understanding how context, and an individual's orientation to social and cultural constructs, informs their behavior (Goodwin & Heritage, 1990). Constructs are instantiated contextually and referenced based on shared background of the actors (Goodwin & Goodwin, 1992). For example, an individual's construct for "fun" is shaped based on cultural constructs, the current context and their shared history with whom they are interacting. This emphasis on context and the individual's constructs make the work of this field particularly relevant to the design of systems that accommodate user perspectives. Throughout this dissertation I will draw on this definition of constructs; a shared, socially structured concept that draws on individuals' background and current context to facilitates discourse.

Given this definition of constructs, their relation to context-aware computing becomes clearer. Specifically, if we assume that constructs are the representation of shared, social understandings, then the interaction of individuals that gives rise to the social structure is of particular interest. How social facts emerge in particular settings is an area of research that has been explored in the work of Lucy Suchman (Suchman, 1987). Suchman's work has sought to understand how social facts develop in work environments and how they are propagated. Her work emphasizes the importance of designing to account for the situated actions of users over plans as a model for human-computer interaction. Emphasizing the importance of situated actions and knowledge, she argues, helps computing systems to avoid the challenges of encoding "common sense" into the plans that computing systems operate upon. The difference is well illustrated when she states:

> "A background assumption, in other words, is generated by the activity of accounting for an action, when the sense of the action is called into question, but there is no particular reason to believe that the assumption actually characterizes the actor's mental state prior to the act. In this respect, the 'world taken for granted' denotes not a mental state, but something outside of our heads that, precisely because it is non-problematically there, we do not need to think about. By the same token, in whatever ways we do find action to be problematical, the world is there to be consulted should we choose to do so. Similarly, we can assume the

> intelligibility of our actions, and as long as the others with whom we interact present no evidence of failing to understand us, we do not need to explain ourselves, yet the grounds and significance of our actions can be explicated endlessly. The situation of action is thus an inexhaustibly rich resource, and the enormous problems of specification that arise in cognitive science's theorizing about intelligible action have less to do with action, than with the project of substituting definite procedures for vague plans, and representations of the situation of action, for action's actual circumstances*"* (Suchman, 1987, p. 33).

Suchman's position that that intelligent actors make sense of behavior by drawing on the knowledge that exists intersubjectively in the world is viewed as a primary fixture of Third Wave HCI.

The position that knowledge exists in the world and is leveraged in our interpersonal interaction to make sense of behavior is not unique to Third Wave HCI. The relationship between phenomenology and ubiquitous, or context-aware, computing was established when ubiquitous computing was first articulated by Mark Weiser (Weiser, 1991). During a keynote talk at the symposium on User Interface Software and Technology Weiser discussed the importance of tacit knowledge in shaping understanding (Weiser, 1994). More recently, Dourish has outlined the theoretical traditions in context-aware computing and notes that in the phenomenological literature, "social facts are emergent properties of interactions, not pre-given or absolute but negotiated, contested and subject to continual processes of interpretation and reinterpretation" (Dourish, 2004).

Exploring the relationship between constructs and context may seem like a purely academic endeavor, but it is an important consideration that should be articulated to position my research. To understand user's perspective, it is critical that we acknowledge that our world—especially our social world—results from a consensus of interpretation. For example, if we sought to build a system that would detect when people were dancing, then we must have an idea of what constitutes dancing. The set of features that correspond to the activity of dancing changes dramatically between social groups and circumstances. These features vary so much that a dancing detection system would be unusable depending on what the user assumed constituted dancing. While some would argue that certain activities are dancing, and others are not, the goal of this work is not to prescribe meaning; this work does not assume that there is a "correct" way to view the world. Instead, the goal of this work is to improve our understanding of how to design interfaces that enable users to communicate their diverse and idiosyncratic perspectives.

1.3 Research Purpose and Thesis Statements

My primary argument is that detecting and responding to context relies on a deeper understanding of the constructs that users' draw on when interacting with context-aware systems. As context-aware systems are asked to infer increasingly subjective, socially situated phenomenon we need to reexamine how context is established and enable users to communicate with the system how the interpretation of the system resonates with their understanding of the context or constructs to which the system is reacting. This is particularly true when the context of interest is affect. In the case of affect, happiness is a construct that often relates to certain physiological indicators, but it is fundamentally subjective. Similarly, if a system were to provide recommendations for "fun things to do," this too is fundamentally subjective. "Fun" is dependent on an individual's constructs for fun, as well as their context. Rather than trying to determine what fun should be, a system should enable the user to express what is fun to them in the given context.

In my dissertation, I explore how the user's perspective is constrained or accommodated by context-aware systems. Furthermore, I design, develop and evaluate interfaces that allow users to negotiate meaning with the system to improve how systems relate to the user. Thus, I am explore the design, development, and evaluation of a strong concept (Höök & Löwgren, 2012) that I refer to as Relating Interfaces. The goal of this dissertation is to investigate, develop, and evaluate interfaces that allow users to easily communicate their perspective to technology. This work expands upon current modes of end-user programming and context-aware recommenders to improve how technology responds to—and ultimately relates to—the user. In my dissertation, I evaluated the following thesis:

> *By affording the communication of constructs, users will feel that Relating Interfaces 1) understand their perspective, 2) provide recommendations that better reflect their perspective than recommendations from non-relating interfaces, and 3) provide recommendations that better accommodate their context.*

1.4 Dissertation Outline

Below is a summary of how this dissertation is structured:

1. Chapter 2 reviews the current literature on context-aware recommendations systems with special emphasis on applications that draw on the emotional state of the user. This chapter contributes an overview of work on context-aware recommendations and the opportunities that exist for future work in this space.
2. Chapter 3 provides insight into how the designers of context-aware systems consider context in their work. I report the findings of 11 in-depth interviews with designers of context-aware systems. This work provides a unique perspective on context-aware systems because of the role that the designer plays in shaping how technology accommodates the user, and thus how technology will be able to relate to users. In these interviews, I followed the design of one system from initial concept to final design chronologically using the artifacts created by the designers as conversational anchors. The interviews and the artifacts were analyzed using methods that draw on conversation analysis and ethnomethodology. Through analyzing of these interviews, I developed an understanding of how their view of context influences what the system detects and how it responds. The work contributes a framework for the design of context-aware systems.
3. Chapter 4 builds of the insights gained from the research in chapter 3 to contribute design guidelines for how to better accommodate context-specific recommendations. From these insights, I developed a relating interface that allows users to convey their perspectives on what music is appropriate for different contexts. I then conducted a lab study with 10 dyads of participants. The participants used a custom chat client to discuss what music would be appropriate for two different contexts: a party and a quite evening with friends. As the participants chatted, the application parsed their conversation for keywords and then visualized those words as a mood board. After the participants had finished chatting the system then used the keyword visualized in the mood board to create music playlist, which they rated for enjoyment, familiarity and appropriateness. Form this evaluation and an analysis of the interviews we discovered distinct strategies the participants used to create context-specific recommendations.
4. Chapter 5 discusses a study in which we evaluated a series of interfaces based on the strategies employed by users to create context-specific playlists in chapter 4. This chapter provides an explanation and rationale for three interfaces we designed and evaluated. To evaluate the interfaces, we created three narrated videos of the interfaces in use. We then

asked 60 participants on the crowd sourcing website Mechanical Turk to rate the interfaces for how effective they would be at providing appropriate and enjoyable recommendations. The findings are also discussed as well as the design implications.

5. Chapter 6 builds on the findings from chapter 5 by developing and evaluating an application inspired by the most highly rated interface from chapter 6. The chapter provides an overview of the design, development and evaluation of a system that provides context-aware music recommendation. In this study, 12 participants used the system for approximately three weeks. During the first two weeks, we logged the music to which the participants listened. Every two hours, the system would check if the participant had listened to music and if they had they were sent a text message with a link to a survey asking them to report their mood, location and a label for their activity while listening to the music we had logged. After two weeks, we created a set of playlists for each participant based on their survey data. We created one playlist for each label they had reported that had been used on more than four survey responses. This resulted in each participant receiving between four and eight playlists with names that corresponded to labels they provided for their activities during the previous two weeks (e.g. *Studying*, *Chilling*, *Working Out*, *Driving)*. For the final week of the study, the participants were asked to report their mood and location and then select a playlist to listen to several times a day. After selecting a playlist, the system randomly assigns them to one of two interfaces. The first interface (the treatment Relating Interface), allowed them to adjust the weighting for songs that would be included in their playlists. In the second (control) condition they would see a preview of which songs would be used to generate their playlist, but that was not interactive. After either adjusting the weights or seeing the preview for the songs, they were then given a playlist of six songs which they were asked to rate for appropriateness and enjoyment. At the end of the week, we concluded the study with an interview with the participants. Through an analysis of the playlist rating and the exit interviews we found that participants enjoyed using the Relating Interface and felt that having the opportunity to adjust how the playlist was created improved their enjoyment of the playlists.
6. Chapter 7 summarizes and discusses the findings of my dissertation. I also include a discussion of the implications for future work.

Table 1.1. Overview of the goals and methods in this dissertation.

Chapter	Goals	Methods	Hypotheses
3	3-1. Determine how designers of context-aware systems approach context in their work.	Thematic analysis of interviews with 11 designers of context-aware systems.	Understanding how designers view context will improve the design of context-aware systems.
4	4-1. Validate a conversational Relating Interface. 4-2. Understand how individuals discuss context-specific music recommendations.	In-lab evaluation of the system with 10 dyads. 1. A statistical analysis of the participants' ratings of the playlists generated. 2. An analysis of the participants' conversation and interviews.	H1. Affording the communication of constructs will improve participants' ratings of recommended music. H2. Visualizing the information used to create music recommendations will improve the participants' rating of the music.
5	5-1. Develop design guidelines for interfaces to communicate contextually appropriate music recommendations.	An online evaluation of three interfaces with 60 participants on Mechanical Turk. The ratings were then analyzed.	Participants' ratings will lead to an improved sense of which interface best accommodate communicating what music is contextually appropriate.
6	6-1. Validate a Relating Interface in a naturalistic setting. 6-2. Determine if Relating Interfaces improve context-aware music recommendations	1. An approximately three-week field deployment of a Relating Interface with 12 participants. 2. A statistical analysis of the participants' song ratings.	H1. Participants will feel that music is more enjoyable and appropriate when using a Relating Interface. H2. Recommendations from a Relating Interface will be rated as more enjoyable and more appropriate.

Chapter 2. BACKGROUND AND RELATED WORK

In this chapter, I introduce and review the relevant background and related work for this dissertation. As was mentioned in the introduction, the focus on this work is on improving users' ability to communicate their meaning with context-aware systems. Much of this is motivated by the idea that HCI has begun to focus more heavily on the personal relationship we have with technology. This chapter provides additional evidence to support the validity and merits of this claim. I will then introduce the role context and context-aware computing, plays in this transition. This will include an emphasis on the role of affect as a component of context. Finally, I will situate the concept of Relating Interface within the framework of context-aware computing.

2.1 HCI and Our Relationship with Computers

The premise of my dissertation is that HCI has begun to transition to a model of computing that encourages a focus on our relationship with computers instead of their ergonomics or cognitive efficiency as has historically been the case. In this chapter I will outline some of the trends that have encouraged this transition and review the relevant background that provides perspective on the HCI paradigms and their position in the current HCI landscape. To do this requires a historical perspective on HCI; so, I will briefly review the relevant technological trends and how they influence the different paradigms.

2.1.1 *HCI Technology Trends and Research Paradigms*

HCI is still a relatively young field, but despite its brief duration it has seen several distinct generations arise in accordance with technological trends. The first generation was ushered in the when mainframe computers became widely available in the 1960s. While mainframe computers existed as early as the 1930s, the first professional operators of mainframes were not present until the release of the IBM System 360 in 1965 (Grudin, 2012). By the end of the 1960s a transition away from mainframe computers began. Mainframe computers were very powerful resources, but they were large, expensive, and difficult to operate. As a result, one computer was used by many people and only for very specific tasks.

In 1968, Douglas Engelbart famously demonstrated the possibility of personal computing and teleconferencing in the so called "mother of all demos". By the early 1980s several personal computers were widely available including the Xerox Star, IBM PC, and the Apple Lisa. This was widely regarded as the second generation of computing; personal computing. Unlike the previous generation, in the second generation of computing each individual had one computer that they interacted with directly, thereby changing the focus away from specialized operators to expecting office employees to have computer literacy as a core work skill. The combination of the personal computer with the Internet led to the one-to-many ratio of computers in which individuals could use their personal computer to connect to any number of computers located around the world. With personal computers widely available, and internet connectivity common and inexpensive, socializing online became more widespread.

In the early 1990s the availability of the Internet and the declining cost of computers led to the new vision of computing pioneered by researchers at IBM. Principal among them was Mark Weiser, who suggested a transition to a model of computing where multiple networked computers of various sizes would be leveraged by users in their daily lives. He argued that computers would become so ubiquitous that they would disappear into our built environments, just like plumbing or electricity. He called this third generation of computing Ubiquitous Computing (Weiser, 1993). Gregory Abowd has suggested that we are currently transitioning into a fourth generation of computing, that he referrers to as Collective Computing (Abowd, 2016). This fourth generation builds off Ubiquitous Computing but blurs the line between what is human and computational.

Parallel to each of these generations of computing is a paradigm motivated by a central metaphor. According to Agre's theory of generative metaphors in technical work, each technical field has a central generative metaphor that structures what are considered valid modes of inquiry (Agre, 1997). In the first generation of HCI, Human Factors researchers considered the ergonomics of interacting with computers and conceptualized HCI as a man-machine coupling. This paradigm's motivating metaphor suggested research questions that focused on efficiency of use. The personal computers of the second generation required more easily comprehensible computers, which led to a concept of the human information processing metaphor (Harrison et al., 2007). This paradigm emphasized the fit between human cognition and machine operations. Terms like 'cognitive engineering' became popularized by researchers, such as Don Norman, who applied classical cognitive science theory to address questions around 'user satisfaction functions' (Grudin, 2005). The third paradigm draws on a metaphor of interaction as phenomenologically situated (Harrison et al., 2007). This suggested modes of inquiry that emphasize the experiential quality of interaction, and a focus on meaning and meaning making.

While overlap exists between the methods and approaches suggested by each paradigm, it is clear that questions around efficiency are far more suited when tasks and goals can be clearly articulated. Similarly, questions around meaning and meaning making become far more relevant when the topic at hand is nebulas and relies on interpretation. My argument is that this is especially true for context-aware computing because, unlike the first and second computing generations, the objects of interest exist in the physical world instead of the digital world. When I access information in a database, that database has a complete 'concept' of that information; it is up to designers and engineers to create a system that represents that information in a way that best fits my understanding and the goals that I hope to accomplish with that information. When the domain of interest is removed from the internal functions of a computer, it becomes imperative that the computer can respond to my understanding of the world. This distinction suggests that third paradigm of computing is the most appropriate for the research in this dissertation.

With that said, I do feel that the sharp distinction that is drawn between phenomenological and cognitive science is overstated. In particular, the literature on third paradigm HCI (Harrison et al., 2007; Sellen et al., 2009) presents cognition as strictly a phenomenon of the brain, which prevents it from being a useful lens for investigating users' embodied action in the world. In my work, I draw on phenomenological literature as well as literature form the theory of distributed cognition that focuses on the whole environment and how individuals interact within that environment (Hutchins, 1995). Distributed cognition seeks to understand the organization of cognitive systems. Unlike traditional theories of cognition, it views the scope of what is considered cognitive as extending beyond the brain to include interactions between people and with resources and materials in the environment (Hollan, Hutchins, & Kirsh, 2000). This focus is useful because it highlights the role that users, computers and their environments play in creating meaning. I see this view as aligning and contributing to current literature in third paradigm HCI and providing additional tools for understanding how individuals act and reason while interacting with computing systems.

2.1.2 *Context in HCI*

Weiser's work on Ubiquitous Computing was a key source of inspiration for the third generation of computing; it offered a vision of the invisible interface in which computers actively detect and respond to users without demanding the user's attention (Weiser, 1994). This vision of computing was an exciting break from the dominant computing paradigm of the 90s that placed the computer in the foreground and necessitated the users' full attention. Weiser's vision was also novel because he positioned it as a phenomenological computing paradigm (Weiser, 1991). During a keynote talk at the symposium on User Interface Software and Technology, Weiser discussed the importance of phenomenological and tacit knowledge in shaping understanding and suggested that ubiquitous computing should draw on these traditions (Weiser, 1994).

After Weiser's passing, the domain of Ubiquitous Computing largely moved away from the phenomenological philosophical tradition. Instead, the field, and their views on context, have been largely positivist. This view of context is represented by Schilit et al. who define context as "where you are, who you are with, and what resources are nearby" and "lighting, noise level, network connectivity, communication costs, communication bandwidth, and even the social situation" (Schilit, Adams, & Want, 1994). Dey expanded the idea of context to include "any information that can be used to characterize the situation of an entity" (Dey, 2001). From this definition we can see that this is perhaps not as much a philosophical position, but a pragmatic position. The descriptions above frames context as a phenomenon that systems can detect and respond to in order to accomplish a goal. Characterizing the situation of an entity is valuable when we know how a system should respond to that specific characterization. However, as was discussed above, as HCI has moved to accommodate less goal-oriented tasks (e.g. maintaining a work-life balance, providing a relaxing ambiance for a dinner party), the specific goals are harder pin point.

An additional challenge to the positivist perspective on context arises from how context is interpreted and how this matches the lived experience the systems detects (Salvador & Anderson, 2003). To understand individuals lived experience it is critical that we acknowledge that our world—especially our social world—results from a consensus of interpretation. For example, if we sought to build a system that would detect when people were dancing, then we must have an idea of what constitutes dancing. The set of features that correspond to the activity of dancing changes dramatically between social groups and circumstances. These features vary so much that a dancing detection system would be unusable depending on what the user assumed constituted dancing. As computers continue to become more prevalent in our lives, they will more often be expected to detect behavior, environments and activities that require interpretation. Therefore, it is necessary they their interpretations is consistent with the lived experience of the individual who utilizes them.

2.1.3 *Affect as Context*

In many ways affect can be viewed as a component of context—albeit, a rather large and important component of context. As such, the field of affective computing has struggled with many of the same challenges as have been encountered defining context, but in ways that are perhaps more pronounced. Within the field of Affective Computing two camps have emerged: the information and the interaction camp. The information camp is dominated by the work of Picard, who draws on psychological or physiological models of affect (Picard, 2000). The interaction camp advocates for a view of affect which draws on social and cultural notions of emotion rather than the psychological or physiological definitions used in the information camp (Boehner, DePaula, Dourish, & Sengers, 2007).

One reason why Picard's work is considered informational is that it discusses affect as a phenomenon that can be abstracted from the individual and expressed as information about that individual. For example, her work only discusses models of emotion that are either discrete and continuous. She states, "Moreover, the question of whether to try to represent emotions with discrete categories or continuous dimension can be considered a choice, as each representation has advantages in different applications" (Picard, 2000, p. 169). Discrete models are those that argue emotion is a discrete, set of finite states. Most early research in emotion points to Darwin for their inspiration who argued for a finite set of emotions and for their evolutionary benefit (Darwin & Darwin, 1872). The discrete emotion tradition was continued by Tomkins who argued for the existence of 8 emotions, which he argued were the expression of underlying biological factors (Tomkins, 1984). The argument for basic emotions was continued, perhaps most famously by, Paul Ekman (Ekman, 1992). In *An Argument for Basic Emotions* he sets out to provide evidence for basic emotional states analyzing the evolutionary basis for facial expressions. For example, he notes that raised eyebrows often characterize the emotion of surprise, and this could be to allow more light to enter the eye, thereby giving the individual experiencing the emotion the opportunity to better analyze the object causing the surprise. Continuous models of emotion suggest that rather than emotion being a set of states, emotion is the expression of some set of dimensions. While there is some variation, all dimensional models adhere to a proposition that neuropsychological and physiological differentiation underlies the approach-avoidance tendencies that lead to positive or negative emotional states (Scherer, 2000). The number of dimensions varies between models, but usually consists of valence and an intensity (Russell, 1980; Thayer, 1978).

One implicit assumption these models make is that the there is no difference between the expression of the emotion and the emotion. For instance, Russell's Circumplex model is based on a dimensional scaling of adjectives sorted into categories by participants' in his study. According to his model, the dimensions arise from the consensus achieved by the participant's sorting of the responses. However, this in no way accounts for the relationship between the internal feeling of the word used in the sorting task and its position in the dimensional space that arises from the task. In dimensional models these two states are the same, thereby reducing the experience of emotion to merely what is reported.

The proponents of the interaction perspective of affective computing argue that this reductionist approach limits the expression of affect to only what can be reported (Boehner et al., 2005). Furthermore, relying on informational approaches to describe affect misses the larger issue that the validity of the model should represent the actual nature and complexity of emotions. Relying on a model for emotion that is inaccurate or incomplete would essential be relying on caricature of emotions, and would unreliable in real world settings. Boehner argues that the HCI community was making a mistake by applying the same informational models to affective computing that it had in prior computing paradigms (Boehner et al., 2005). Early work in HCI often relied on simplified, goal-oriented models of user behavior (Maguire, 2001). Boehner argued that the same was happening in affective computing by relying on overly simplified models of emotion.

Boehner's position seems sound, but her position has limitations. The work the interactional camp draws on is based on language-based studies of emotion. These methods are fundamentally limited because they rely on the participant's awareness of their emotional state, but people are not necessarily conscious of our emotional states. Renowned neuroscientist Antonio Damasio suggests that people are unconscious of a great deal of our emotions. He state "there is no evidence that we are conscious of *all* our feelings, and much to suggest that we are not" (Damasio, 2000). Another issue that is caused by relying solely on self-report emotional states raises is how well the user can correctly appraise their current emotional state (Cacioppo, 2000). Research has shown that sunny weather tends to elevate people's moods, whereas rainy days can be depressing (Clore & Ortony, 2008). The weather thus serves as a source of affective reactions that might not always be correctly attributed to their true source. Consequently, people tend to respond that they are in better moods and more satisfied with their life on sunny days than on rainy ones. However, when participants were first asked directly about the weather, thereby linking their feelings to their true cause, the effect of weather on judgments of life satisfaction disappears. Making weather salient as the true cause of the feelings makes its impact on the appraisal of mood go away. It did not change the feelings, but it did change their judgments about life satisfaction. Therefore, individuals are informed by their affect, even when the affect is self-produced. (Clore & Palmer, 2009; Wyer, Clore, & Isbell, 1999). This suggests that individuals may act in accord with emotions that they are not consciously aware of, and attribute to other stimuli, making it difficult to assess their actual emotional state.

From this we can see that, while affect is an important factor in shaping context it presents serious difficulties to designers and researchers. Both the information perspective and the interactional perspective rely on assumptions of what is the correct way to interpret affect and both have short comings. Detecting affect as a form of information is insufficient for providing a valid account of an individual's emotional state and instead only represents the emotional state they hope to present. Asking individuals to report their own emotional state also poses complications due to misattribution or unawareness on their part. Instead of relying on strict interpretations of affect, it seems reasonable to allow users to use it as a resource in how they interact with system rather than prescribing a meaning associated with their emotion. In this dissertation, I rely on participants to report their emotions, but avoid tying their self-report data to external notions of how emotions should be used. Instead, I treat emotions as a personal tool for demarcating the significance relative to each participant for their context.

2.2 Related Work

Now that I've reviewed the relevant background literature and theory behind my work, I will position the contribution of my research in the HCI landscape. Fundamentally, I see my work as a contribution to the design of context-aware systems. As such, I will begin by discussing work that explores the design of context-aware systems and the user's notion of context. While the domain of this dissertation is the design of context-aware systems, a large portion of this dissertation focuses on the design and evaluation of interfaces for music recommendation systems. The reason for this is primarily methodological. A great deal of excellent work in the field of HCI utilizes qualitative modes of inquiry to build theory and gain insight. In this work, I chose to develop theory and then test that theory through quantitative evaluations to provide evidence on the efficacy of the system. Therefore, I found it necessary to provide a measure of success that could align with the theoretical position that draws on the situated experiences of the individuals. Music therefore offers an ideal mechanism for evaluation. Music is highly personal, contextually relevant, and is a component of how activity and behavior is structured (Denora, 1999). Also, because music can be assessed in a short period, it is easier for participants to respond to it in a variety of settings. In the remainder of this section I will outline the related work from context-aware computing, affective computing, and recommender systems that respond to context with an emphasis on music recommender systems.

2.2.1 *Context-Aware System Design*

This dissertation is situated within a body of literature examining the role of context in design. To support the design of context-aware systems, researchers have taken several approaches. Of particular relevance to this work are studies of context-aware systems that explore how meaning is created between users and the system. Mediation is a technique that has been explored in prior work (Dey & Mankoff, 2005; Dey, Mankoff, Abowd, & Carter, 2002). The work on mediation attempts to improve ambiguity by providing a GUI so that system can respond to inputs from the users. While this work is an important contribution, it sees the issue of ambiguity as arising from detection error not from an inability of system to adopt the user's perspective. Research on ambiguity is continued by Lim et al. who explored issues relating to intelligibility (B. Y. Lim & Dey, 2011, 2013; B. Y. Lim, Dey, & Avrahami, 2009). His work carefully varied how information was presented to determine when it is considered helpful versus overwhelming. While my work does not explore intelligibility directly, this dissertation contributes to work on intelligibility and mediation by exploring how context can be negotiated by users and the interfaces that are most effective to this end. The role of ambiguity in context-aware systems has also been explored as a resource for design with the Home Health Horoscope (Gaver, Sengers, Kerridge, & Kaye, 2007). This system provided ambiguous messages based on sensor readings from various sensors placed throughout the home. Their work provides insight into how users reflect on ambiguity, which our work builds from by providing design insights into how context is viewed by designers and users.

Another important topic of work is determining a shared view of context. Several approaches have been taken to understand context from the user's perspective. Indexicality has been used to enable users to address their environment in prior work (Kjeldskov & Paay, 2006; Rantanen, 2010). Similar to the work on mediating interfaces, this work was limited to physical and assumed that misinterpretation for the physical world was the result of imperfect detection. Furthermore, it does not consider the users' social world. Additionally, a great deal of research in end-user programming creates mappings between the detected environment and how the system should respond (Dey, Hamid, Beckmann, Li, & Hsu, 2004; Sohn & Dey, 2003; Truong, Huang, & Abowd, 2004). This dissertation builds on this work by accommodating a multi-faceted representation of users' understanding rather than the mapping predetermined by researchers or developers.

Context has been leveraged to improve recommendations as well. While the notion of context varies across domains (Adomavicius & Tuzhilin, 2015), some definition align more clearly with the definition of context described above. Many system leverage location as a key component to improve recommendations for activities that could be of interest to the user (Zheng, Zheng, Xie, & Yang, 2010) or information on nearby geo-coded wikipedia articles (Simon & Fröhlich, 2007). Other work has sought to provide recommendations to users for places to meet by using overlap in GPS coordinates (Khetarpaul & Gupta, 2012). Social networks have also been explored as a method for improving recommendations (Gonzalo-Alonso, De Juan, Garcí-A-Hortelano, & Iglesias, 2009), however, their formulation of context was inferred from online presence rather than their behavior in the world or their physical environments. The work in this dissertation contributes to the work on context-aware recommendations by providing concrete insights into how users discuss their views of context and how this influences the relevancy of recommendations.

2.2.2 *Music Recommender Systems*

Determining what music is desirable or appropriate for a given context has been a topic of extensive interest in the field of HCI. One of the earliest examples was MusicFX (McCarthy &

Anagnost, 1998), a group based recommender system, which determines a playlist for gym members based on overlap in their music preferences. More recent work has sought to create playlists for groups based on music they vote for (O'Hara et al., 2004; Sørensen & Lagerl, 2012; Sprague, Wu, & Tory, 2008), their emotional valence (Bauer, Jansen, & Cirimele, 2011), their activities (Liu & Reimer, 2008), or prior listening habits and copresence (Crossen, Budzik, & Hammond, 2002). While prior work provides useful insights into how music can be recommended to groups of individuals, it does not provide insight into how groups determine the conditional relevance of music for a social context. This dissertation builds on this by providing details of how individuals leverage music as a resource in the creation of a social context.

The social role of music has also been studied in work that considers music's role in sharing and collaboration. For example, Brown et al.'s (Brown, Sellen, & Geelhoed, 2001) study of music sharing highlighted the role of sharing as a social practice and explored alternatives to illegal file sharing services. More recent studies of how music is used in social activities have provided insights into people's interactions surrounding music, including how it is shared and experienced (Leong & Wright, 2013). Voida's work on practices of music sharing importantly illustrates the role of impression management through music (Voida, Grinter, Ducheneaut, Edwards, & Newman, 2005), while Sease and McDonald (Sease & McDonald, 2009) contribute to this by revealing the role that intimacy and proximity play in how music is used in impression management. This work contributes to these studies by providing insights into how individuals establish the relevance of music for a social context, thereby influencing what experience individuals hope to create.

In addition to user studies on the social role of music, several music sharing systems have been developed and evaluated to explore how users share music and listen collaboratively. This exploration of music sharing has included systems for sharing between collocated (Seeburger, Foth, & Tjondronegoro, 2012) and non-collocated users (Baumann, Jung, Bassoli, & Wisniowski, 2007; Liu & Reimer, 2008). Through the use of a cultural probe, Pocketsong, Kirk et al. (Kirk et al., 2016) provide a detailed account of the practices of listening and sharing social experiences through the affordances of mobile music applications. While the practices of sharing and experiencing music are important areas of research, our research contributes to prior work by providing insight into how users determine what criteria are important when music is used to create the tone or interactional framework of a social event.

2.3 SUMMARY

Current computing trends suggest a world where technology is more proactively aware of the user's context in a variety of ways. This includes not just the users' physical context, but their emotional and social context as well. This offers several important opportunities to designers and developers, but only if systems can detect the world in a way that matches the lived experience of the user. The ability of computers to sense and infer context is likely to improve, however the role that computers play in users' lives suggests a mode of interaction that do not assume a ground truth but rather a consensus of interpretation. While previous work has sought to establish this consensus through mediation and intelligibility, this dissertation builds on prior work by contributing insight into how designers and users view and respond to their context. It also provides evidence of the validity of the concept by presenting a quantitative evaluation of interfaces that draw on these insights.

Chapter 3. WHAT DESIGNERS TALK ABOUT WHEN THEY TALK ABOUT CONTEXT

The nature of context and its role in the design and development of technology has been a topic of much debate in the field of human-computer interaction (HCI) (Dey & Abowd, 2000; Dourish, 2004; Greenberg, 2001). Now, in an era where context-aware systems are regarded as commonplace (Abowd, 2012), context's role in design is no longer just a topic of debate for researchers—it is a daily concern for practitioners. The perspective of designers is of particular interest in the process of creating context-aware systems because of the role they play in determining how a technology will be situated in the world and therefore what constitutes the system's context. However, how designers understanding of context is reflected in their work is a topic of research that is conspicuously absent from the literature.[2]

While much has been written about the nature of context, much of the literature has focused on a pragmatic approach to what can be detected by computers (Dey, 2001; Schilit, Hilbert, & Trevor, 2002) or on a theoretical exploration of the nature of context (Dourish, 2004). Rather than explore the nature of context, in this chapter, I investigate designers' views on context. Recent work has helped to clarify the value that taking a practitioner's perspective can provide to the larger research domain (Goodman & Wakkary, 2011; Stolterman, 2008). To take the practitioner's perspective means that we learn to understand context not as a neutral, objective phenomenon based on technology or on theoretical models, but as a construct that reflects the views of designers. In this chapter, I aim to determine designers' understanding of context by examining the artifacts they create and methods they utilize in the creation of context-aware systems. Put differently, this chapter provides an examination of what designers talk about when they talk about context.

To achieve a better understanding of context-aware design practices, I conducted interviews with eleven designers of context-aware systems. During the interviews, designers provided us with the artifacts produced in the design of one context-aware system they had created (see Figure 3.1). They then walked us through the design process detailing the role of the artifacts and methods in which they engaged. This allowed us to follow the design process from initial concept to a finished product, all from the designers' perspectives.

[2] Portions of this Chapter first appeared as Jared S. Bauer, Mark W. Newman, Julie A. Kientz 2014. "What Designers Talk About When They Talk About Context," Human–Computer Interaction.

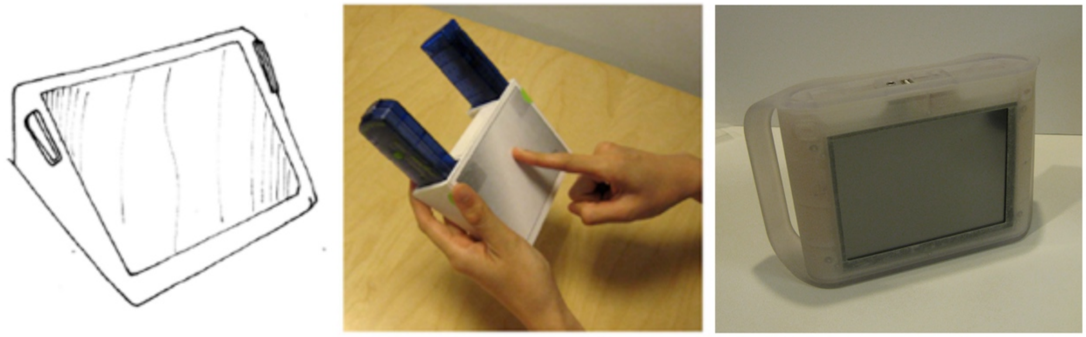


Figure 3.1. Artifacts from interview with Participant 4.

This chapter contributes a detailed description of the design processes of context-aware systems from the designers' perspective. Our analysis revealed a set of five concerns designers need to address to produce a coherent system. These concerns are: *users*, *context*, *form*, *interaction*, and *implementation*. It further revealed a pattern of activity by which the designers' understanding of context influenced how they *framed* the design space, *filtered* the design space for possible solutions, *encoded* the contextual components of the system into a vocabulary, used this encoded vocabulary to *unify* a solution, and then *evaluated* the solution in terms of the codes. Additionally, rather than acting as a static concept through this process, my collaborators and I found that the designers' concepts of context adapted in light of possible solutions and newfound constraints. This forced designers to revisit these stages multiple times to evaluate various solutions.

This chapter contributes to my dissertation by providing empirical insights into designers' views on context. One of the primary goals of my dissertation is to improve how designers accommodate context in their work. By contributing insights into current practices, this work helps provide a designer focused premise of Relating Interfaces as a strong concept (Höök & Löwgren, 2012).

3.1 Related Work

This work draws on empirical studies of designers and previous work attempting to characterize context. The title of this article is an intentional reference to Paul Dourish's work investigating context and its role in design (Dourish, 2004). Dourish's work outlines the value of viewing context from a phenomenological perspective in which context arises from our interaction in the world, drawing on everyday, cultural, common-sense understandings of the nature of the social world. He contrasts this perspective with positivist accounts of context where it is viewed as a set of attributes of the world that can be objectively observed and enumerated. The positivist view of context is represented by Schilit et al. who define context as "where you are, who you are with, and what resources are nearby" and "lighting, noise level, network connectivity, communication costs, communication bandwidth, and even the social situation" (Schilit et al., 1994). Dey expanded the idea of context to include "any information that can be used to characterize the situation of an entity" (Dey, 2001). The purpose of this chapter is not to discuss the merits of these positions. Each is valuable, but more importantly, they are formulated as academic positions on the nature of context and do not necessarily represent how context is viewed by practitioners designing context-aware systems. These positions on the nature of context represent the experience of HCI researchers designing and developing context-aware systems; however, this work contributes to prior work by providing empirical evidence to enrich the HCI literature on these positions.

3.1.1 *Theoretical Lens*

In this investigation of design practice, my collaborators and I draw on Schön's (Schön, 1992) theory of *design worlds* to inform our understanding of the conceptual space in which design work is conducted. Schön argues that through designers' perceptions of actual or virtual worlds, they create the objects and relationships with which they interact and determine what exists in the design world. These design worlds are abstract spaces in which designers create and evaluate objects and relations as they work to create an optimal design. Rather than investigating particular objects in the designers' world, my collaborators and I explore how the designer creates the relationships among objects. Our contention is that the designer's formulation and representation of these relationships necessarily influences the type of technology that is suited to exist in that design world. This means that when a designer formulates how a new design will respond to context in their design world, it is based on an understanding of context that will guide how that design is able to relate to the world. As Schön notes, design worlds may be unique to the designer or shared across a community of practice. This suggests that explicating these worlds, and the types of relationships inherent to them, could help to establish a common ground to reason about modes of context-dependent interaction.

To explicate these worlds, my collaborators and I draw on Charles Goodwin's (Goodwin, 1994) theory of *professional vision*. Professional vision is the term Goodwin developed to describe the practices used by members of a profession to shape a domain of scrutiny. For example, an anthropologist and a farmer impose different meanings on to the same substance (or "domain") – e.g., soil. Their analyses of the domain rely on different assumptions, methods of analysis, and systems of scrutiny. Goodwin argues that these practices create the knowledge that forms the theories, artifacts, and expertise that are distinctive to any professional domain. According to Goodwin, professional vision relies on three practices in any domain:

> *"1) Coding schemes used to transform the materials being attended to in a specific setting into the objects of knowledge; 2) highlighting, making specific phenomena in a complex perceptual field salient by marking them in some fashion; and 3) the production and articulation of material representations"* (Goodwin, 1994, p. 2).

By analyzing the practices in the domain of design for context-aware systems, we can begin to understand the ways designers understand context as a domain of scrutiny.

3.1.2 *Context Within Design*

Our study is situated within a body of literature examining the role of context in design, with special attention paid to the practices of designers. The seminal work on Contextual Design by Beyer & Holtzblatt (Beyer & Holtzblatt, 1999) has argued for the importance of understanding the context of use for design and outlined methods for understanding context from the perspective of users. While understanding the context of use is an important aspect of design, our work differs by focusing on systems that are proactively aware of what this context might be. Prior work has explored how Information Architecture (Morville & Rosenfeld, 2008) design methods and artifacts account for context (Bauer, Newman, & Kientz, 2014). While this work is useful in characterizing how context is manifested in design for Information Architecture, the present article builds on it by providing a comprehensive account of the design process of context-aware systems. Methods such as Experience Prototyping (Buchenau & Suri, 2000) seek to enable designers or other stakeholders to engage with the imagined uses of a system, including understanding the role of context. Similarly, Davidoff et al. present the Speed Dating method, that they argue allows designers to rapidly explore application concepts and their interactions and contextual dimensions(Davidoff, Lee, Dey, & Zimmerman, 2007). While each of these methods creates compelling ways to explore context in design work, they do not provide insight into designers' views on context.

Previous research has supported the design and development of context-aware applications, including application toolkits (Dey, Abowd, & Salber, 2001) and infrastructure support (Hong & Landay, 2001) to facilitate the rapid development of context-aware applications. These efforts have produced a number of insights into the technical requirements for supporting context-awareness and the potential for easing the burden of development, but they have been primarily aimed at software developers who have a different set of concerns, practices, and skills than designers.

To support the design of context-aware systems, researchers have taken several approaches. Prior work has sought to use *design patterns* to support design for ubiquitous computing systems (E. S. Chung et al., 2004; Landay & Borriello, 2003). While designers did find such an approach useful, the generated patterns were based on a review of the research literature instead of being informed by observed design practices. Dow *et al.* conducted a series of interviews with designers to investigate design practices for context-aware systems (Dow, Saponas, Li, & Landay, 2006). This work revealed the importance of storytelling for depicting context in design. However, Dow's work largely focused on issues influencing the development of tools to support ubiquitous computing designers rather than the designers' understanding of context.

3.2 STUDY METHODS

We conducted 11 video-recorded interviews with designers who had worked on recent projects that gave "special consideration to the users' context," as quoted from our recruitment email. My collaborators and I chose a wide framing of context to capture projects that designers themselves described as being context-driven. Additionally, I discussed the projects with the designers prior to the interview to ensure they possessed characteristics that meet Schilit et al.'s definition of "context-aware" (e.g. software that "adapts according to the location of use, the collection of nearby people, hosts, and accessible devices, as well as to changes to such things over time" (Schilit et al., 1994). Following prior studies of designers (Dow et al., 2006; Newman & Landay, 2000), I focused each interview on a single project on which the designer(s) worked in the recent past.

The interviews focused on design artifacts and methods, how the artifacts were used, and with whom they were used to communicate in the design process. My collaborators and I define an *artifact* as any tangible or digital document, device, object, or file that was created by the designer to help them in the design process (Blackwell, Whitley, Good, & Petre, 2001; Norman, 1991). This definition includes a variety of items produced by designers, such as wireframes, architecture diagrams, mockups, interactive prototypes, physical models, written documents, paper or dry erase board sketches, videos, and photographs. Artifacts were used as memory triggers and interview foci, but they were also reviewed closely in conjunction with the transcripts for the role they played in each specific design project. This allowed us to understand how the artifacts and the methods were applied to the concerns of each project. Prior to each interview, I asked designers to provide me with an inventory of artifacts produced during the selected project and provide copies, photographs, or scans of the artifacts wherever possible. Because our study was retrospective in nature, our request for the artifact inventory included several prompts and examples of different types of artifacts to help designers recall a greater number of details regarding the project under study.

3.2.1 *Participants and Recruitment Criteria*

We recruited via messages posted to several interaction design mailing lists, through personal contacts working in industry, and by searching for publicly shared design projects or products that seemed appropriate for the study and directly contacting the designers. For our study, my collaborators and I defined designers as individuals who engaged in an iterative process to explore a problem space and produce a product or service that addresses the problem they explored. My collaborators and I view this approach as contrasting with scientific, artistic, or engineering approaches. Our view on the distinction between design and science or art draws on Cross's definition of 'designerly' ways of knowing (Cross, 1982, 2001). Unlike art, Cross argues that design aims to provide a solution to a problem. Cross also argues that science focuses on solving a specific, well-formed problem, whereas designers address ill-defined problems. Similarly, my collaborators and I view engineering as setting out to produce a solution to a known problem rather than working to create a solution to an ill-defined problem. While recruiting, my collaborators and I sought designers that worked in either academic or industry settings.

Because our definition for designers was specific, once I had conducted 10 interviews, my collaborators and I decided to exclude two of the interviews from further analysis. My collaborators and I chose not to analyze these interviews because the process discussed focused too heavily on engineering a solution and therefore were not useful for understanding design practices. Because my collaborators and I chose to exclude these interviews, I returned to recruiting and conducted three additional interviews to ensure that my collaborators and I reached a point of data saturation (Bowen, 2008; Lincoln, Egon G. Guba, & Guba, 1985). Thus, this chapter presents findings from 11 interviews, which included a total of 14 designers, as two of the interviews (1 and 8) were conducted with multiple designers (see Table 3.1).

Table 3.1. Participant experience as designers and the projects discussed.

ID		Years as Designer	System Description	Platform
1	**A** **B** **C**	1 2 0	Location-aware smartphone application for locating restaurants.	Mobile Phone, iOS
2		12	Wearable location and activity sensing smartphone application for logging physical exercise.	Windows mobile, custom sensor suite
3		10	Small tangible tile capable of recognizing gestures and proximity of other tiles.	Custom-built tangible interactive device
4		3	Suite of sensors and portable educational tool for high school science students.	Custom-built touch screen device with a suite of sensors
5		9	On-body activity and location sensing device and smartphone interface.	Custom hardware and mobile phones
6		5	Tangible device used to promote mindfulness of power consumption.	Custom tangible device implemented with micro-processors
7		4	Interactive TV and ambient interface for socializing through the television.	Custom hardware and commercial televisions
8	**A** **B**	4.5 6	Location aware smartphone game.	Mobile phone, iOS
9		10	Location-based desktop and smartphone app for healthy lifestyle recommendations.	Web Applications, iOS
10		5	Location-based smartphone application for managing time.	Mobile phone, iOS
11		4	Mobile application for sharing online shopping experiences.	Mobile phone, iOS

3.2.2 *Coding and Analysis*

We conducted a two-stage analysis of the interviews. In the first stage, my collaborators and I conducted a content analysis of the interviews by coding each interview using the Text

Analysis Markup System (TAMS) Analyzer software [3]. To identify common concerns and themes in the interviews, we created a code list and then coded the transcripts of the interviews. To generate a list of codes, the first author watched portions of each interview and developed lists of potential codes. These codes were then reviewed to see how consistently they appeared in other interviews as themes. Members of the research team then watched an interview to check for consistency in the application of codes. This led us to refine and remove several codes and then reapply codes to the interviews. The final code set was influenced by Lim *et al.*'s notion of prototypes as filters in design work (Y.-K. Lim, Stolterman, & Tenenberg, 2008). Per Lim *et al.*, when viewed as filters, design artifacts are used to investigate, or filter, the qualities in which designers are interested, without distorting the understanding of the system as a whole. My collaborators and I expand the categories outlined by Lim *et al*. to account for some of the phenomena we felt these designers were 'filtering.' We added *context* and *implementation* to their initial list of filter categories to describe the phenomena that we felt the designers were trying to explore with their artifacts. This initial stage of coding allowed us to identify the concerns designers addressed in their work.

After identifying the salient concerns addressed by designers, I returned to the artifacts and interviews to explore how the designers view of context was manifest in their artifacts and practices and how their view on context influenced their work. To explore these questions, I applied professional vision (Goodwin, 1994) as a theoretical lens to the design artifacts and the transcripts of the interviews. My collaborators and I looked for examples where designers attempted to discuss and account for context in the design of the system. Because the focus of this work is to understand context from the perspective of the designers, we attempted to remain neutral about what context could mean. Therefore, rather than analyzing the interviews by looking for examples of what we believed context to be based on a literature review or personal intuition, we instead looked for representations or practices that are not accounted for in conventional computing interfaces (Hutchins & Hollan, 1985) or that relied on implicit interaction (Schmidt, 2000). Additionally, we looked for instances where the designer specifically discussed context.

[3] http://tamsys.sourceforge.net/

3.3 FINDINGS

In this section, we outline the findings from the analysis of the interviews and artifacts in two ways. First, we begin by detailing the concerns that designers addressed across the interviews, where I briefly discuss the concerns we identified. I then provide a description of a design process to demonstrate how designers addressed these concerns. Finally, my collaborators and I turn our attention to the patterns of activities that characterized how designers we interviewed moved through the design process and illustrate these with relevant examples.

3.3.1 *Five Common Design Concerns*

The analysis of the interviews revealed that the designers addressed five common concerns. While specific design questions were unique to each project, we found the following higher-level concerns to be useful in characterizing what aspects of the system the designers sought to understand. The five concerns we encountered were: *users, context, form, interaction,* and *implementation* (see Table 3.2). By identifying these concerns, I aim to provide insight into the emphases that the designers explore through their artifacts and methods. I am not asserting that there is a 'correct' amount of emphasis to be placed on any given concern, but it was clear that the extent to which these concerns were visited affected the outcome of project. Additionally, providing a vocabulary of concerns helps to add clarity to the utility of various methods and practices.

Table 3.2. Common concerns that context-aware designers address.

Concern	Example Questions
Users	• What experience and knowledge do the intended users possess? • What are the intended users' goals, preferences, mental models, and current behaviors?
Context	• In what physical environments will the product be used? • In what concurrent activities will users be engaged at the time of use?
Form	• What physical form should the system take? • How large, heavy, rugged, attractive can/should the devices that make up the system be?
Interaction	• How will the system receive input from the users and the environment? • How will the system communicate with users?

Implementation	• What are the requirements for hardware and software in terms of capability, performance, power consumption, etc.? • What is it possible for the implementation to achieve in terms of these factors?

3.3.1.1 Design Concerns in Practice: Interview 8 Background

To illustrate how the concerns above were represented in the practices of context-aware designers, I will now detail the design process from Interview 8. This interview is a useful example for two reasons. First, their work resulted in a successful and well-received game, so their process can be called successful in some objective way. Secondly, this team had several years of prior design experience (see Table 3.1), but no experience in context-aware design. This interview is helpful in illustrating the unique challenges that designing to account for context can create.

Interview 8 was conducted with two designers about their work developing a new game for smartphones that used the location of the users as a factor of the game (see Figure 3.1). The designers were part of a team of four masters students that had received funds to develop the game from a competition financed by a local entrepreneur. The funds paid for them to work on the project over the summer between their first and second year in their masters program. The team had four members, and while they all were involved in the design and development of the application, our interview was with the two who had focused primarily on its design.

Figure 3.2. The finished application designed by the participants from Interview 8.

Both designers had a background in web design. Designer 8A noted having 3 years of experience working on web design and front-end web development. His undergraduate education was in business and entrepreneurship, but he was eager to have the design and development skills to "start a project and get [it] running." Similarly, 8B had a degree in Business Information Technology. Prior to returning to school for his Masters degree, he had worked doing web design and development. He commented that his web design experience focused on "evaluation of potential [web] applications. And…adding features that might be easy to add via a bit of code." Despite having a background in web design, they felt that mobile, location-aware design was going to be an increasingly important area in design. They saw their work on the project as having value beyond the immediate product as it also gave them the chance to develop their professional skills and build their portfolios in a burgeoning design domain.

Winning the competition was viewed as prestigious, and previous winners had turned their ideas into successful startups. However, accepting the grant money was not without risks. Because their work on the project occurred during the summer between their first and second years as Master students, it prevented them from having internships. Receiving this grant and forgoing internships meant their work on the project would represent a large portion of their professional portfolio as they entered the job market. The opportunity that the grant for this project afforded then was significant, but because they were working on a new topic with no mentors over a summer break, they knew there was the very real possibility that the project could fail. Also, because the technical domain was reasonably new, they were not certain what would or would not be feasible in the allotted time.

Another factor that affected their work was the relationship of the team. The two designers interviewed for the study knew each other prior to their work on this project; they were classmates and part of the same social circle at school. However, they had only met the two developers they were working with to form a group with the purpose of entering the competition. Thus, the design and development portions of the group did not have a history of working together. This resulted in the members of the group having difficulty selling each other on their ideas. The group's entry into the funding competition proposed that they would build a location aware game for the iPhone, but other than that, they had decided on very few specifics. Because the nature of the game they intended to build was undecided, and they were newly acquainted, the power structure of the group was of particular importance.

It became apparent during our interview that they had no clear group leader and instead relied on achieving consensus when making decisions. While decision-making was diplomatic, both designers mentioned how challenging it was to persuade other group members to take particular directions with the game they were developing. They stated that they would nervously practice their pitches before group meetings and that meetings would be long and contentious. While group members seemed respectful of each other, the lack of familiarity, and thus trust, required the designers to argue their case persuasively. This resulted in the team's work being delayed through much of the summer as they decided on a direction to take the game. At times, it became clear that the challenges imposed by the group dynamics influenced the ordering and creation of the artifacts in their design work.

3.3.1.2 Design Concerns in Practice: Interview 8 Overview

During their work, they created seven artifacts in all (see Table 3.3). As per our interview, the work is discussed chronologically using the artifacts as foci. This will help to establish the process by which these artifacts were created and what purpose they served. Doing so can illustrate how the designers addressed various concerns and how successful their artifacts and methods were in this process.

Table 3.3. Artifacts and the concerns they address from Interview 8.

Concern	Artifact 1: Table sketches	Artifact 2: User research findings	Artifact 3: Whiteboard sketches	Artifact 4: Lo-fi wireframes	Artifact 5: Demo application (prototype)	Artifact 6: Whiteboard sketches	Artifact 7: Final prototype
User		●			●		●
Context	●	●		●	●		●
Form					●		
Interaction	●	●	●	●	●	●	●
Implementation	●		●	●	●		

The designers began their work by sketching ideas on a glass table in their office (**Artifact 1**). During this stage, they iterated over a variety of ideas for games. Each of these ideas ultimately revolved around a set of location-dependent user interactions. This process was very influential in forming their thinking of how the user's location—or *context*—and *interaction* could be used as components of a game. They also spent much of their time discussing how they could *implement* a system that would respond to these two elements. When discussing this stage in their design process one commented:

Participant 8a: *"So we knew we wanted to do a game—a game using the GPS capabilities of the iPhone that makes the physical world the game board. That was the kind of assumption that started all this reasoning. We had a few ideas in mind in the very early stages."*

Once they had developed a few ideas for games, they conducted informal interviews with 6 or 7 peers to get feedback. The designers kept a hand-written synopsis of the interviews on a white board in their offices throughout the design process and referenced it regularly (**Artifact 2**). These findings distilled their interviews into a set of high-level themes about potential *us*er attitudes toward *context*-aware games and what forms of *interaction* in the game might be compelling. Both designers noted that these notes were regularly referenced during debates about the direction they should take with the game. This was part of the reason why they kept the notes on the dry erase board through the duration of the summer. The visibility of these notes was useful in keeping the direction of the project aligned with their findings. With these findings in mind, they began sketching versions of the interface on a whiteboard (**Artifact 3**). Externalizing their ideas as sketches gave them an opportunity to discuss how some of these games could be played and how they might implement their ideas, given that they were planning on developing an application for the iPhone. These sketches allowed them to understand how they could *implement* the desired *interaction* of the system.

With their idea for the system externalized in sketches, they developed more detailed wireframes of the interface to determine how the user's *interaction* and location would affect the game (**Artifact 4**). During this process, they again returned to the sketches on the glass table (**Artifact 1**) to visualize the role of location in the game (see**Figure** 3.3**)**. The locations on the glass table were hypothetical, but were used to understand how changes in location, or *context*, would impact the interface and thus the users' *interaction* with system and how they might *implement* the wireframe as a functional application. When commenting on this process one designer stated:

Participant 8b: *"[The glass table] is actually a little better than the white boards, 'cause you can kind of hover over it like you would a map; and given that we were working on a map-based thing…at one point I think we even considered putting—like, we have a map that we considered putting underneath it just to kind reinforce the idea that whatever we're doing is based on the geography of the real world."*

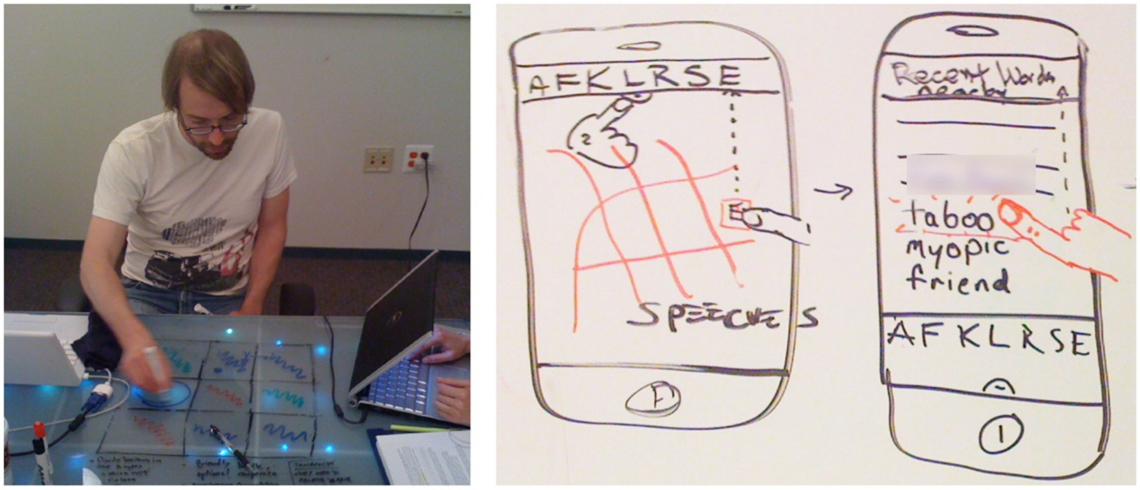


Figure 3.3. Artifacts from interview 8.

After they had completed the wireframes, they developed a demo application (**Artifact 5**) that they gave to a group of peers to try out during a street fair. The initial game mechanic was a type of scavenger hunt that the participants were asked to play as they wandered through the street fair. This form of prototyping allowed them to gain *user* feedback about the core *interaction* of the system in *context* running on the iPhone (and thus in the correct *form*). One of the outcomes of this process was the realization that they had focused on the *implementation* too heavily, and this resulted in them having modes of interaction that did not necessarily result in a fun game.

> **Participant 8a:** *"I mean, we have this cool technology, that you can pick up stuff from a map, you can drop stuff in a map based on your geolocation, you can pass stuff between users, there is user-to-user interaction going on, from iPhone to iPhone, cool, I send it to you…but what—we don't have…I guess the problem is where is the story, the compelling story, where is the narrative of the game, from that part, and we felt that we were not accomplishing anything."*

This finding from their prototyping led them to reevaluate and redesign the narrative of the game. This was complicated by the fact that the summer was coming to its end and they needed to produce a finished product. During a debate about which direction to take the game, designer 8A introduced the idea of a scavenger hunt game. The objective of the game was to collect pieces of an image scattered in the physical world and then reassemble the image. They hoped to monetize the game by allowing sponsors to suggest images, which could be assembled and returned for prizes. He explained the game by suggesting that a local bar could give a discount on a beer to people that collected and returned a re-assembled image of a beer. As he spelled out the word "Beer" on the dry erase board the group had an epiphany. They decided to create a word game like scrabble, but with the letters out in the world waiting to be picked up. This idea was appealing because it was consistent with their initial goal of having a location-aware game, but also aligned nicely with the user feedback they had gathered. On a more pragmatic note, this game could be developed without having to create complicated graphics, which had been a sticking point on prior ideas since neither of the designers came from a graphic design background.

With an idea for the game finally settled, the remainder of the work for the project proceeded rapidly. Having a defined direction was helpful, but the rapidly approaching end of summer gave them additional motivation. They began redesigning the application by sketching ideas for the interface. This allowed them to determine the user's *interaction* with the system (**Artifact 6**). After several rounds of sketching, they finalized the idea in a higher fidelity prototype and conducted informal testing to evaluate what *users* thought about the *interaction* in *context* (**Artifact 7**).

Their finished application received primarily positive reviews on Apple's App Store and game review websites; it also had a large and sustained user-base, so by many definitions their work was a success. However, to arrive at the final design required significant changes to the application after their first round of prototyping. The issue they encountered with the initial application did not arise from a lack of thought about the game play mechanics; during the summer, they vetted numerous ideas and discussed games at length before deciding to create the scavenger hunt game. So, what caused the issues with their initial concept for the game? By looking at the concerns and the order they addressed them, it becomes clear that from the earliest stages in the design process, they were already concentrating on how they would implement the system. Three of their first four artifacts included some reflection on how they would implement their ideas. As a result, when they finally created a prototype to gain insight on the users' interaction with the system in context, they were dissatisfied with the result. It is hardly a new finding that focusing on implementation early in the design process can preclude possible design directions. However, it was not necessarily their focus on implementation that was the sole issue; we believe a more significant issue was that the designers did not produce artifacts that allowed them to simultaneously triangulate on the interplay between *user, context,* and *interaction* concerns. As a result, they focused on implementing the system to enable them to understand these three concerns in conjunction instead of creating artifacts (e.g., prototypes) to explore these concerns directly.

The difficulty in finding ways to explore the user's interaction with the system in context was a common theme across several other interviews. In fact, this concern was mentioned in five of the eleven interviews (Interviews 1, 6, 7, 8, and 10). In regards to this challenge, the designers from Interview 1 noted:

> **Participant 1a:** *"We were developing things just to see what it would look like. It was part of the design in that it was just to fool around and prove to ourselves that we could actually make something. See what our limitations were... "*
>
> **Participant 1b:** *"but also see how people could interact with it"*

While this concern was common, it was not uniformly distributed. It was observed that the more senior designers, particularly those with experience doing prior work on context-aware systems (e.g. Designers 2 and 5), produced artifacts that more evenly addressed these concerns. More importantly, they did so without having to implement functional prototypes. Because of this, I believe that designers learn to orient toward these concerns and gravitate toward practices that enable them to understand issues that arise at the intersection of these concerns. The following section discusses in greater depth the processes designers employed to explore these concerns and the role their view of context played in shaping this process.

3.3.2 *Accounting for Context in Design*

The previous discussed the concerns that must be balanced while designing context-aware systems. With these concerns in mind, we turn our attention to the process by which designers investigate these concerns during their work. As was seen above in the work of the designers from Interview 8, balancing concerns is fundamental to this process, but how these concerns are balanced, and what role the designer's view on context plays in this process, has yet to be explored. In this section, I outline four practices that my collaborators and I observed designers engaging in during the design of a context-aware system. Specifically, this section details the practices designers followed to *frame* a design space, *encode* the relevant criteria against to create a context-aware system, *unify* possible solutions within that design space, and then *evaluate* possible solutions. Each of these practices is expanded upon below.

3.3.2.1 Framing the Design Space

Designers' views of context frame the space in which the system they are designing will exist. This process of framing begins when designers articulate what they hope to accomplish in their work and how they might accomplish it. In this sense, the process of framing a design space is largely constructive, but what it constructs is a space bound by the designer's view of context. Gaver's research on design workbooks (Gaver, 2011) suggests that design creates the spaces in which it operates. Our findings are aligned with this, but in addition to creating the space in which the design operates, the designers' notion of context creates a way to frame that space such that certain notions of context are clearly contained while others are not. This process can be seen unfolding by analyzing what is, and is not, represented in the designer's artifacts. A wide range of things could constitute the user's context in a given design space, but by relying on what the designers do represent in their artifacts, notions of context frame the design space in distinctly different ways. An example of the process can be found be found in the work of Designer 5 while designing an on-body device for runners that adds informational markers to locations as the user runs. Designer 5 is the principal designer for a design firm, and she has a wealth of experience in context-aware and other forms of design. Figure 7 shows a sequence of 4 of the 24 frames that made up a storyboard she created at the beginning of her work on the system (see Figure 3.4).

Figure 3.4. Four frames from Participant 5's storyboard.

What the storyboard includes and omits reveals a great deal about what the designer considered to be the relevant features of context. In this artifact, the designer includes the locations (represented by the buildings) and the activity (represented by the illustration of the character running). In the second frame from the left, the designer's annotation shows the text that the system would provide to the runner as they interact with the device to "leave a tag." In the third panel, the designer shows a second location and again uses text to indicate how the system responds to the location and its ability to infer the pace of the runner based on their location. The response of the system also suggests that it is aware of the route the user has taken. Designer 5 explained that the storyboard was useful to help her understand the modes of interaction. She stated:

> **Designer 5:** *"This was a very, very rough storyboard of how something might work in context; so, trying to identify what are the points that there needs to be visual or audio feedback, and what are the points there needs to be input—voice input, basically."*

Her quote and the storyboard suggest that the designer views the context as arising from the user's activity and the locations where that activity occurs. This can be seen in the third frame in which the system responds to her activity at a given location. If either the activity or the location were absent, the caption she included would make no sense. The feedback the system provides in the caption is clearly based on both the runner's pace and the locations where the tags had been placed. Furthermore, we can see that the locations the system responds to are not coordinates, but physical locations demarcated by the building in frames one and three. By depicting the system responding to the user's pace at physical locations by tagging the locations, the designer demonstrates that she views the context as emerging from the user's interaction with the world based on their location and activity. Context, in this example, could have been just the location or just the activity, but by viewing the context as involving the interaction of the two components, the designer frames the design space in a specific way. Furthermore, by focusing on the physical locations and making the system respond to these locations, she further frames a space to explore in her work. The design space she subsequently explores still contains numerous potential solutions, but her concept of context frames that space such that a successful design solution will include physical locations and the user's activity at those locations.

In Interview 9, an alternate formulation and representation of context imposes a different framing of a design space. Interview 9 was conducted with a senior designer at a major international design firm who also had considerable design experience, including experience with the design of context-aware systems. We discussed his work on an application to encourage healthy lifestyles. He used a "screen storyboard" in his design work on a system to recommend healthy lifestyle options to commuters (see

Figure 3.5). This artifact provides an interesting contrast to how context was represented by Designer 9 in two ways. First, instead of representing a user, it only represents the commuter's route, thereby reframing the idea of context away from an embodied experience to a general trajectory that the user could follow on their commute. Second, his storyboard has the locations of interest to the user represented as "dots." By representing the context as dots corresponding to fixed coordinates, as opposed to buildings or another representation of location, the designer frames the context as being solely derived from the user's location.

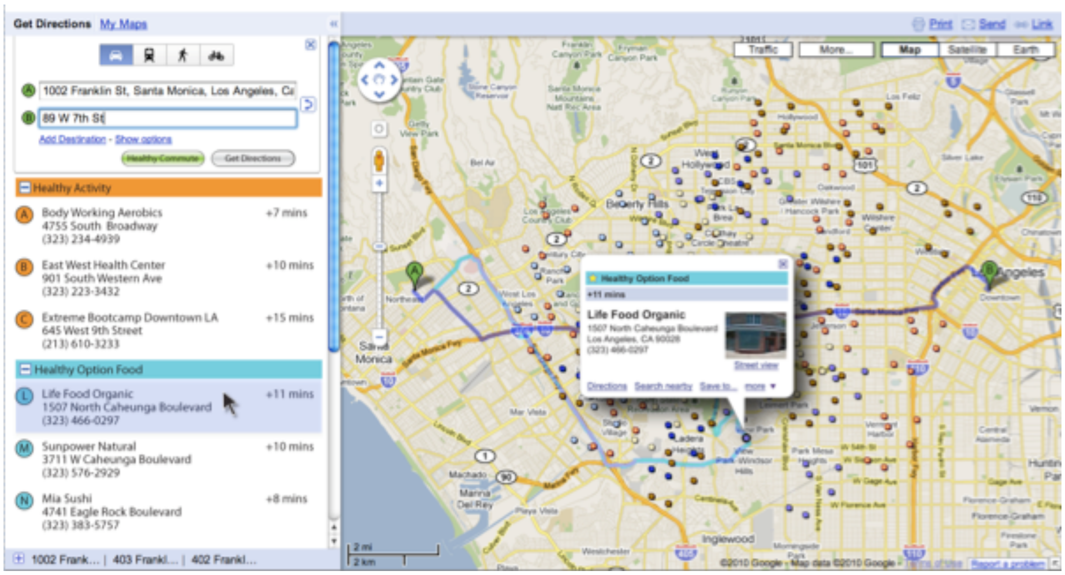


Figure 3.5. Figure 8: A "screen storyboard" created by Participant 9.

The systems discussed in Interviews 5 and 9 each contained a representation of the user's location and a route that the user traversed. However, these examples clearly contain differences in how these components were viewed. Designer 5 viewed the location as physical spaces and the route as something that resulted from the user's activity. Designer 9 viewed locations as a set of fixed coordinates and the route as a path that linked the user's current location to their home or work. In both cases, the way the designer framed the location imposed different constraints on the space their work inhabited and therefore the types of questions they would need to answer in further stages of the design process. Neither of these notions of context is 'right,' however, these differing views will shape the space they explore to find a satisfactory solution.

3.3.2.2 Encoding and Evaluating Context

In addition to framing the space in which designers work, designers' view of context plays an important role in establishing which solutions in that space are optimal or even feasible. This process has two steps, first *encoding* their concept of context and then *evaluating* various solutions in terms of the codes. To clarify, the process of encoding is not a deliberate activity by designers; it occurs incidentally when designers articulate concepts as they work. By articulating their concepts relating to context, designers instantiate sets of codes for the context-dependent components of the system. According to Goodwin, coding is the systematic practice by which professionals of a domain transform the world into the categories and events relevant to the work of their profession (Goodwin, 1994). In this instance, designers encode the notion of context in a specific way allowing them to communicate the ephemeral aspects of the context to peers and collaborators. It also establishes the basis on which the system will be evaluated.

In the interviews, we found numerous examples of codes helping to reify the designers' concepts for context. By analyzing the appearance of these codes, and how they were manifest in the designer's work, we can see that these concepts evolved through the design process. The location-aware smartphone game from Interview 8 illustrates this point. During the design process, the designers developed the "picking up," "dropping off," and "passing" interaction techniques that were central to their game. Each of the actions relied on the designers' view of how close the users were to the virtual objects or other players. In this example the phrase "passing" establishes a specific relationship between the users, the context (e.g. location and collocation), and their interaction with the application. Such codes further instantiate what they think "near" and "collocated" mean, based on the phone's ability to discern those factors. By using these codes, they transformed a complicated nebulous phenomenon of "picking up" to the distances between the geo-locations of a user and a virtual object.

Designers' concepts of context transform again when the codes begin to be used as criteria against which design objectives are evaluated. The initial articulation of codes, in many cases, draws on a commonsense understanding of the world and our interaction in it. As the work progresses the designers must determine if their work is leading them to a finished product that satisfies their view for the context-dependent components of the system. Designer 2's work on an activity-aware health and fitness application provides an example of this process. The application she was developing would respond to the user's location or fitness activity, such as running, walking or using an elliptical trainer. These different fitness activities were detected by a custom-made sensor suite worn on the user's hip, and were communicated to the user's smartphone. Most people share a commonsense notion of what running is, but the designer must determine if the system responds to their notion of running in terms of the goals for the system. In this sense “running” becomes a code, and more importantly the designer must determine if the system responds appropriately when it thinks the user is running. While discussing the process of designing around the detected activity of the user, Designer 2 noted:

> **Designer 2:** *“And see here, you can see in this one there’s a question that we wrote and there’s a big [collaborating developer’s name] with a question mark so we would bring him over and there’s another [collaborating developer’s name] with a question mark. So [collaborator] was working with [different collaborator] and so when we would hit some ‘Ooh is that even possible?’ We would write it on the board and you can see we start working out the details of how things would get triggered.”*

This process of creating codes to crystalize a concept for context and then evaluate that idea was apparent across all the projects my collaborators and I reviewed; the designers articulated an initial set of codes, which would reify the designer’s notion of context in the artifacts. Articulating context with artifacts allowed them to move their idea from the designer’s world to the physical world. This enabled the designers to see what possible sets of solutions existed in the design space they had framed. As designers winnowed the design space, their notions for what constitutes context also narrowed. Finally, their concept for context was crystalized as a set of constraints that must be satisfied to demonstrate that the system is context-aware. During this process designers realize the shortcomings in their initial concepts for context, for example that they were too imprecise, limited, or infeasible. This resulted in the designer shifting their concept for context and thereby reframing the design space. The process of framing and coding occurred iteratively as designers searched for optimal solutions to enact their ideas.

Concordant with our observations, Cross has suggested that codes translate abstract requirements into concrete objects and that creating codes is a fundamental component of ‘designerly ways of knowing’ (Cross, 1982, 2001). One important distinction in context-aware design is that the designers are obliged to describe the relationship between their concept for the system and the world. While some areas of design in HCI—information visualization or information architecture for example—can happen in the abstract, context-aware design fundamentally requires the designer to articulate and encode the relationship between the system and the world. Because the relationship between the system and the world must be encoded, context-aware designers’ practices are more clearly aligned with Goodwin’s concept of coding (1994). Goodwin argues that coding schemas allow professional practitioners to translate the relevant information in the world into objects of knowledge. He goes on to argue that this process of encoding and communicating these codes are fundamental to the discourse of a profession. It is notable, though, that while designers are compelled to develop codes to reason about context on a project-specific basis, few *shared* codes for context currently exist in the HCI community that are formalized in a consistent way.

3.3.2.3 Context as a Unifying Element

As designers explored various solutions, their concept for context served to unify the various components of the system. Put another way, the way designers encoded context imposed constraints on the system. As was discussed above, context-relevant codes also served to establish evaluation criteria against which the success of various solutions could be judged. Moving toward a solution required creating artifacts that enabled the designer to explore multiple concerns simultaneously. This process usually involved the creation of increasingly sophisticated artifacts to evaluate the design. The fact that artifacts increased in sophistication through the design process is unsurprising (Newman & Landay, 2000). However, the difficulty designers faced in creating artifacts that allowed them to evaluate their ideas was surprising. This resulted in several tradeoffs being made to create artifacts that could implement their ideas. These artifacts generally were constructed toward the end of the design process and provide insight into how the designers' view of context had evolved. Our analysis of these artifacts revealed that context played an important role in aligning the different design concerns. The difficulty inherent in determining and responding to context forced the designers to either winnow the design space to notions of context that were easier to implement or to explore previously unexplored areas of the design space.

One example of how context unified the designer's concept comes from Interview 7. The designer in this interview was a senior researcher in a research and design department of a major international technology firm. In the project we discussed, he was developing a new system that would enable social interaction through television. The context in question they were designing for was the social presence of the user's peers. The system allowed users to watch and comment on shows with their remotely located peers, but communicating others' presence and desire to watch television posed a series of prototyping challenges. Designer 7 explored several techniques to convey the peers' social presence including building a prototype using an Ambient Orb[4].

---

[4] http://www.ambientdevices.com/about/consumer-devices

However, he and his collaborators were ultimately dissatisfied with these options. In regard to exploring possible options, he commented:

> **Designer 7:** *"So we did things like that, and it was largely a search for…the problem was we had these hardware requirements we knew we wanted, like we had this picture of the ideal thing that we could use, and nothing in the market really met that very well.*
>
> **Interviewer:** *"So you developed that prototype to sort of figure out what would be some other options because this system didn't actually satisfy some of the constraints that you had? Was it particularly the Wi-Fi?"*
>
> **Designer 7:** *"That was the main thing that we really needed. We have pretty good control over…the other thing was the, yeah. The Wi-Fi also, in addition to the reception problems, there was also a factor of the update response. So, you know, the…if you just go and buy an ambient orb from Brookstone or something like that, and you plug it in, the updates that you get are going to be within a time window of, I think, something like fifteen minutes or something like that, and we wanted something that was much more responsive than that."*

The "picture of the ideal thing" he refers to clearly entails some encoded notion of context. Essentially, he felt that a change in the social presence of the user's peers could be abrupt. To convey such changes required a way of communicating remote users' presence that would render changes visible in a short period. The Ambient Orb could not communicate this abruptness, thus they had to explore a different solution. Their final design did allow for abrupt changes in user presence, thereby allowing them to align other system aspects with their concept of context. To consider alternatives, they developed a blog of various technologies that could be used to communicate presence. While discussing this approach, Designer 7 stated:

> **Designer 7:** *"We had a list of requirements as to how we wanted these things to function, and I think the blog just showed like a number of different ways in which you could have like connectivity, you could have visibility, and how much information you could convey."*

From this quote, it is clear that the designer was attempting to satisfy a set of concerns and prior prototypes had revealed shortcomings in their initial approach. To achieve a design that unified all the constraints their concept entailed, they reframed their initial idea in a way that allowed them to explore a new form and method of implementation, thereby creating a solution that unified the elements of their notion of context.

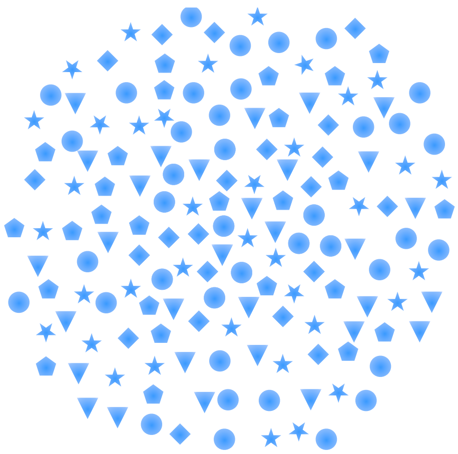
3.4 DISCUSSION

The goal of this work is to understand the practices of design for context-aware systems from the perspective of the designers creating these systems. Now that I have detailed the concerns designers addressed and the process by which those concerns are resolved, I revisit the analysis to clarify how designers deploy these practices to transform their knowledge and views on context into a functional, context-aware system. As I revisit the findings, I discuss their implications for process, practices, and tool support, as well as their limitations.

3.4.1 *Toward a Context-Aware Design Process*

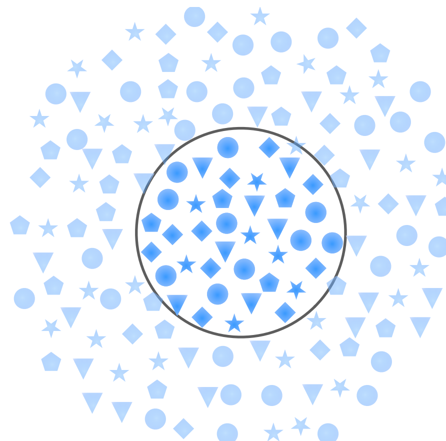
Our findings outline the processes by which the designers engaged with context and the role that context played in this process. Based on these findings, my collaborators and I propose the following process model to characterize the practice of context-aware design (see Figure 3.6).

Design Space
Initially, there is an unlimited set of possibilities for the designer to explore (not all known to the designer).



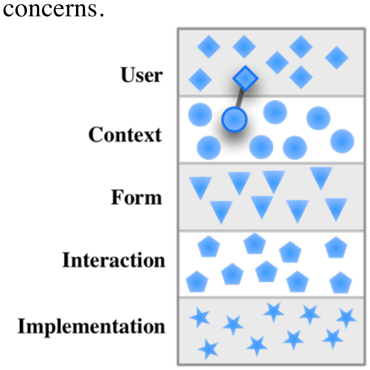
Each design concern is represented here by a symbol (user, context, form, interaction, and implementation).

1. Framing
Designers begin by articulating ideas for the behavior of a system and how it will be situated in the world. This frames the space the designer will explore.



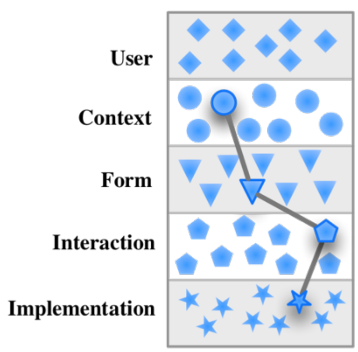
Example: The designer decides to build a fitness application for runners.

2. Encoding
Further elaboration of the concept encodes the behavior of the system as a relationship between the context and the other concerns.



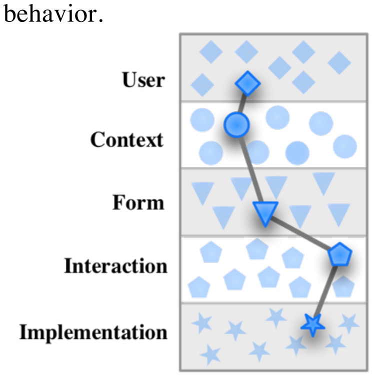
Example: It will detect when the user is running by looking for a change in their location over time.

3. Unifying
As the designer continues to explore the design space, constraints are surfaced that establish specific relationships.



Example: Determining if the user is running (*context*) will require knowing their exact location (*implementation*). An application for runners should be portable (*form*) and largely hands free (*interaction*).

4. Evaluating
The designer will then evaluate various solutions for the system against the codes they have established for the system's behavior.



Example: The designer has found a way to implement the system that satisfies their idea for the user, context, form, interaction, and implementation.

Figure 3.6. Context-aware design process model.

My collaborators and I see this process as having four phases. However, we do not suggest that these phases are visited only once. Rather, we observed that designers revisited these phases multiple times as their work evolved. Furthermore, we do not suggest a strict ordering to which these phases are visited. However, the order they are presented represented the order they are likely visited at least initially. These phases are:

Framing: The designer(s) articulate and explore a concept of context, which imposes a set of limitations on what exist inside and outside of the design space their work inhabits.

Encoding: As the designer(s) discuss the behavior of the system, they begin to instantiate a vocabulary, or codes, to express this behavior. Encoding allows them to discuss the behavior of the system in a way that corresponds to their initial view as they continue to develop the system.

Unifying: As the designer(s) explore the design space, certain possible design solutions are brought to the foreground. These solutions impose additional constraints on the other concerns the designer addresses. As they seek to create a solution that corresponds to their encoded notion of context, ways to satisfy the various concerns will coalesce into a unified solution.

Evaluating: As the process continues, the designer will focus on a solution that satisfies the constraints according to their encoded formulation of context. This allows them to determine when they have arrived at a satisfactory solution.

Perhaps the most important question when looking at the concerns and process model I have outlined is "what good does this do and for whom?" For practitioners, having a defined language to discuss concerns inherent to the work on context-aware design can lead to a more comprehensive approach. In the interviews, my collaborators and I saw that more experienced designers better balanced these concerns. However, if they were to provide practical suggestions to new designers, they might struggle to communicate which concerns are general, which are specific to a given project, and which arise because of focusing on context. By providing a vocabulary of concerns, designers can better communicate how these concerns are addressed. Secondly, the process model contributes a way to view the role of context in influencing design work. For example, the initial formulation of context may be taken for granted if designers fail to identify the role the concept for context plays in framing a design space. Similarly, acknowledging the ways a concept of context unifies a solution serves to draw attention to the formulation of context rather than the solutions which it instantiates. In general, providing a vocabulary to discuss context-aware design, we believe, will help practitioners and academics evaluate the concept that a system embodies.

One valuable exercise when viewing this model is considering what is *not* context. As noted in the methods section, we recruited participants using an intentionally wide framing for context. Despite our initial broad framing, my collaborators and I are not of the opinion that everything *is* context. It is our view that context-aware systems enable modes of interaction or experiences that draw on factors external to the system, such as its context of use, features of its environment, as well as the unfolding interaction with the user. Additionally, the designer of a context-aware system should be able to communicate how those factors impact the behavior of the system. Focusing on contextual factors expands upon traditional concerns in HCI such as the user's tasks, goals, and internal cognitive states, which have been the focus of other design models (e.g., GOMS (Card, Newell, & Moran, 1983) or TCUID (C. Lewis & Rieman, 1993). My collaborators and I are not arguing that this model replaces prior design models. Instead, we believe this model is useful for understanding the additional implications to a design when enabling interaction or experiences that draw on these externalities. From this, it is clear that if a concern does not draw on external factors to shape the user's interaction or experience with a system, then that *is not* necessarily engaging notions of context. Thus, this model is not a way to view design work in terms of context, but rather serves to make explicit assumptions the designer has made about the context of the system.

3.4.2 *Toward Context-Aware Design Practices*

A major theme my collaborators and I encountered in the research was designers' difficulty producing artifacts that allowed them to gain an understanding of how context would impact the user's interaction with the system. To counter this, my collaborators and I saw designers adapt "standard" artifacts (e.g., storyboards, flowcharts, wireframes) to distill the complicated notion of context down to individual components of context needed to express the concept in a manner the artifact affords. For example, this can be seen with the designers in interview 8 when they distilled the context of the system down to simple grid of locations drawn on a table. Certainly, location was an important component of context but, this method did not provide a rich enough understanding of context to completely evaluate the design concept. In contrast, by using Wizard of Oz techniques and rapid prototyping tools (e.g., Arduino[5]), designers were sometimes able to interactionally instantiate some forms of context, which allows the contextualized behavior of the system to be experienced (e.g. Designer 7's work prototyping a social television experience). By adapting existing artifacts, designers can consider the constraints and opportunities of context with other long-standing concerns.

In addition to affecting the content of artifacts and the details of their use within the design process, a focus on context-awareness foregrounds *new design practices* and their attending artifacts. The most notable example of this is Experience Prototyping (Buchenau & Suri, 2000), which recommends a certain approach to the creation of low-fidelity prototypes and/or simulated environments along with a set of techniques for engaging prospective users in contextually-grounded usage experiences. Despite the advantages that Experience Prototyping offers context-aware design, my collaborators and I found few designers employed this practice in the projects we documented. This finding can be interpreted in several ways. One possibility is that knowledge of how and when to apply such techniques is not yet widespread—even though the academic Design community has known of this technique for over a decade. Alternatively, it may again point to differences in how designers view context. Experience Prototyping allows for an embodied, interactionally-instantiated experience of the system. This sharply contrasts other artifacts that designers relied on that use a positivistic representation of context to understand and/or specify the envisioned behavior of the system.

The contrast between which artifacts were chosen and how they were applied sheds light on the tension between phenomenological and positivist views of context. The findings demonstrate that both perspectives exist in contemporary context-aware design practice. Interestingly, both perspectives can be found within individual projects as pursued by individual designers, though admittedly this was less common than seeing one or the other. While my collaborators and I are reluctant to make claims about consistent temporal patterns for activities and artifacts based on our study, our data suggests a provisional alignment of the phenomenological perspective with earlier design stages and of the positivist perspective with later design stages. Such an alignment makes sense because as the system progresses, it becomes increasingly necessary that designers communicate the design in a way that can be implemented,

[5] http://www.arduino.cc/

which suggests an increasing emphasis on views of context that can be more readily captured, modeled, and acted upon by computing systems with discrete inputs and outputs.

3.4.3 *Prescriptions for Context-Aware Design*

Our analysis revealed the importance of generating and communicating codes to the members of the design team and the role that various artifacts play in the design of context-aware systems. This finding draws our attention back to Schön's (Schön, 1992) discussion of design worlds. Schön argued that design worlds may be unique to a designer or shared across a broader community. This suggests that an analysis of the specific codes could help to establish a common vocabulary for context-aware design. Initiatives to facilitate designers' communication around their context-aware design practices could serve to facilitate the emergence of standard codes or coding practices. This in turn could improve the utility in design patterns for context-aware systems.

A question raised by this work is: how *ought* designers consider context within established design process methodologies? For example, this work might be seen as asking how to reconcile efforts to understand, represent, and account for context within the stages of a standard user-centered design process (Moggridge, 2006; Saffer, 2006). From this perspective, this work might seek to prescribe particular practices or representations that are appropriate for considering different types of context in the various stages of needs assessment, design exploration, prototyping, and evaluation. Alternatively, this could be seen as suggesting an alternative process models such as Hartson and Hix's Star lifecycle model (Hartson & Hix, 1989) that emphasize the separation of concerns and encourages alternating one's focus between specific design issues or activities and evaluation of the integrated whole. Indeed, my collaborators and I were struck by the seeming absence of a consistent temporal pattern of activities and/or artifacts when we compared across projects. Some projects focused on technical concerns first (including the feasibility of detecting particular contextual states or events), others sought to understand users' needs (especially with regards to various contexts of use), while others began with concerns about interaction, navigation, and content organization (including the expected behavior of the system in the face of contextual events). The non-linear processes thus observed appeared to align with Moggridge's (2006) suggestion that design should begin by first considering the relevant constraints. He also suggests that the designer ought to move between these concerns in a noncyclical fashion. An open question is whether such flexibility is to be discouraged or embraced and whether standard models of the interaction design process are sufficient when designing context-aware systems.

This work has implications for the development of tools to support context-aware design as well. As noted earlier, application frameworks and toolkits for building context-aware applications (e.g., (Dey et al., 2001; Newman et al., 2010)) can speed up the design life cycle significantly. However, this research highlights the fact that these tools, which focus attention on specifying system behavior, are more appropriate for later stages of design when a positivist perspective may be more fruitful. Low-fidelity prototyping tools like Topiary (Li, Hong, & Landay, 2004) or Activity Designer (Li & Landay, 2008) allow designers to take greater advantage of ambiguity, thereby reducing the need to specify all details of the system before exploring its interactional characteristics. However, even these tools emphasize the specification of screen layouts and state transitions, forcing consideration of the details of discrete contextual inputs and states.

Earlier still in the design process, tools that allow designers to explore interactional context, articulate and test assumptions, and experience alternative designs within different contexts of use will play an important role. ChronoViz (Fouse, Weibel, Hutchins, & Hollan, 2011) is an example of a tool that could be used by designers to collect, organize, and visualize sensor traces of user behavior to gain an understanding of the nature and diversity of contexts relevant to a given application. RePlay (Newman et al., 2010) extends these capabilities by allowing captured sensor data to be fed into application prototypes as they are being developed, thus closing the loop between early and late design and supporting the transition from phenomenological and positivist modes of thinking about context. Wizard of Oz (WOz) tools (e.g., (Dow et al., 2005; Li, Hong, & Landay, 2007; MacIntyre, Gandy, Dow, & Bolter, 2004)) can play an important role in facilitating the consideration of context throughout the design process, though such tools generally require that the designer specify a concrete set of valid inputs that can be simulated during a WOz experiment. Cleverly deployed, however, WOz tools can leave designers with enough flexibility to be able to improvise alternative contextual states and/or system responses, thereby allowing exploration of a system's interactional context in parallel with trying to ascertain its concrete behavior.

3.4.4 *Limitations*

This research relied on artifacts and the designers' recollections of methods. As was discussed above, it is my view that the process of creating artifacts necessarily influences the way context is represented. Relying on the artifacts may have influenced designers to think about the process in terms of the representations of the process, which may create a bias toward a positivist interpretation of context. Practices such as experience prototyping serve to represent the designer's concept of context, but by relying on their memory of the practice, it undoubtedly loses some of the richness that being there would reveal. Despite this limitation, I do feel that the designers were able to discuss their design practices with sufficient detail for the analysis. However, an ethnographic study of context-aware design practice would be a valuable way to explore this topic in future work.

Finally, my collaborators and I sought to investigate the practices of context-aware designers, and, to that end, with whom we were able to conduct interviews limits our view of context-aware design practice. I cannot say how representative the interviews I conducted are, but because of the convergence I began to see by the 11th interview, I suspect that these interviews are broadly representative. Furthermore, our findings aligned with the results of previous work on this area, which again suggests that the practices my collaborators and I observed were representative of the practices at large.

3.5 Conclusion

In this chapter, I have sought to contribute to the understanding of context-aware design by analyzing the artifacts and practices of designers who have worked on projects where context-awareness was a key component to the system's functionality. In doing so this chapter addresses research goal 3-1 and provides insight into how designers of context-aware systems approach context in their work. Our findings suggest designers' view of context is both phenomenological and positivist, and that both views play important roles in the design process. However, the choice of methods and artifacts influences how one's view is manifest in the system that is developed. Designers' views evolve as they seek to satisfy five concerns in their work: *users*, *context*, *form*, *interaction*, and *implementation*. By addressing these concerns designers' understanding of context influences how they *frame* a design space, *encode* the contextual components of the system into a vocabulary, use this encoded vocabulary to *unify* a solution, and then *evaluate* the solution in terms of the codes. The process is not a straightforward march, but relies on creating multiple representations of the context that are evaluated in conjunction with different concerns and by different stakeholders.

This chapter contributes grounded insights on context from the perspective of the designers. These insights helped to inform the design of Relating Interfaces, which are developed and evaluated in subsequent chapters. Furthermore, it contributes a vocabulary and framework for discussing the design and evaluation of these systems. By investigating designers' views of context, this work provides valuable insight into how to improve system's ability to communicate meaning in the design of context-aware systems.

Chapter 4. REFLEKTOR: EXPLORING SOCIAL, SITUATED MUSIC RECOMMENDATIONS

Music is fundamental to the social lives of people around the world (Davidson, 2004). Whether it is Pomp and Circumstance played at graduation ceremonies, hymns sung by a congregation, or pop hits enjoyed by friends at a club, music is nearly always present when people interact socially. But music isn't merely a component of an occasion; music also helps to establish the tone and conduct expected for a given context (Crozier, 1997; DeNora, 2005; Shepherd & Wicke, 1997). Music's role in structuring social interaction has been studied extensively by ethnomusicologists have argued music acts as soundtracks for social action or as a, "*framework for the organization of social agency, a framework for how people perceive (consciously or subconsciously) potential avenues of conduct*" (DeNora, 2000, p. 17). Music's ability to structure social conduct and norms extends beyond individuals; music acts a medium through which individuals create an event (Crozier, 1997; Denora, 1999). As music becomes more readily available, the role music plays in structuring our social interaction is likely to increase. In fact, Hargreaves and North have argue that due to technological trends that have vastly increased the portability, availability, and reproducibility of music in the 21st century, the focus of music in everyday life has transitioned from being primarily cognitive or emotional to social (Hargreaves & North, 1999). [6]

Recent work in the field of HCI helps to shed light on the social aspects and practices of music listening and sharing (Brown et al., 2001; Leong & Wright, 2013; Voida et al., 2005), but how individuals negotiate the relationship between their context and the music for that context is a topic previously overlooked. One of the challenges in exploring how individuals determine context-specific music recommendations stems from the difficulty in determining how to study context in general. In the field of HCI, prior work has defined context as the information that can be used to characterize the situation of an entity (Dey, 2001; Dey & Abowd, 2000). More recently, context has been defined as being an emergent property of individuals' interaction instead of the information that characterizes it (Dourish, 2004). Conceptions of context as something that emerges from interaction nicely align with ethnographic work suggesting that music is

[6] The work in this chapter was done in collaboration with Julie Kientz and Aubury Jellenek. Jellenek was primarily responsible for interface design and development on the Reflektor application and facilitated the in the lab studies.

fundamental to structuring social conduct. By presuming that music contributes to the structure of social conduct, and that context emerges from this interaction, we can see the role music plays in co-creating context. One method often utilized in ethnographic work for understanding how context is established comes from conversation analysis (Goodwin & Heritage, 1990). Therefore, drawing on the conversations of individuals while they discuss context-specific music recommendations is an ideal way to better understand how music is used to create social context.

To explore how individuals use music to create social context, my collaborators and I developed and evaluated an application called Reflektor. Reflektor is a chat-client that parses the users' conversation for key words and phrases, which it uses to create an interactive visualization and music playlists. Rather than trying to infer what users want and then present suggestions, our system sets out to accommodate how users discuss music to provide insights into how users create music playlists for a given context. By exploring how individuals discuss context-specific music with Reflektor, this work contributes an in-depth analysis of how users determine what music is appropriate to shape a given context. Additionally, this work surfaces mismatches between how context is grounded and the metadata recommender systems draw on to create music recommendations. This chapter provides insights into the strategies that participants used to establish common ground around music and context. Therefore, this chapter helps to establish the principles that accommodate the users' views when developing Relating Interfaces.

4.1 METHODS

In this section we provide an overview of how we explored the practices of creating context-specific music recommendations. Because we explored this topic with Reflektor, we begin with an overview of the system used by the participants during the study. We then detail the study design and our approach toward analyzing the participants' chat conversations.

4.1.1 *Reflektor System Description*

Reflektor operates as one system but has three distinct components (see Figure 4.1). The first is a *chat client* where the participants can chat with one another, share, and vote on the keywords from their conversation. The second component is a *server* where a text parsing module determines the keywords and phrases based on the participants' conversation and then uses those keywords to produces music recommendations. The final component is a dynamic *visualization* of the keywords from the participants' conversation. The participants interacted with the system using a smartphone running the chat client, and the visualization appeared on a wall-mounted display in the room where the study took place. Each of these components is explained in detail below.

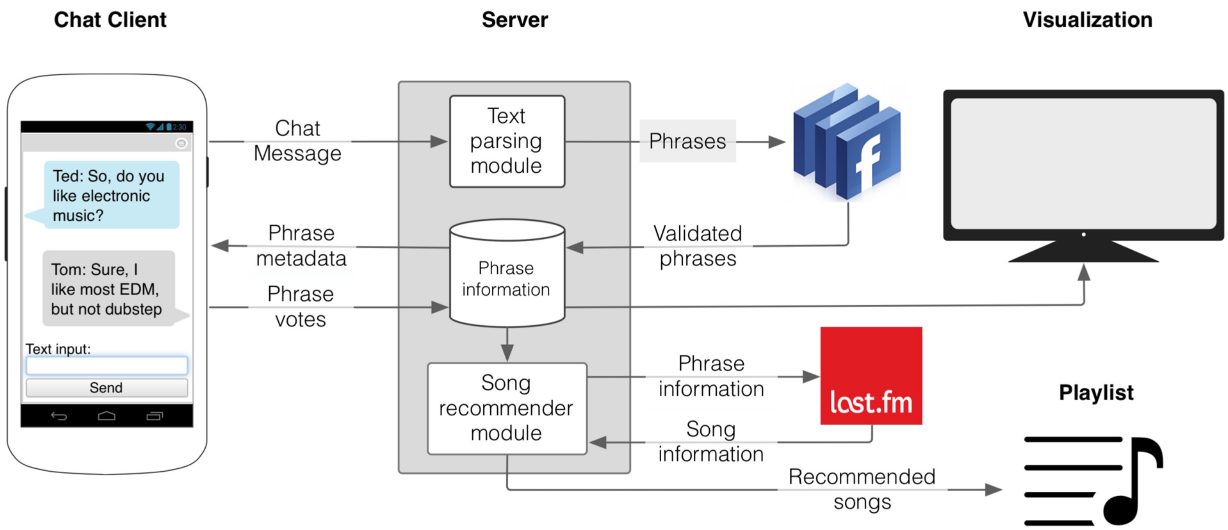


Figure 4.1. System diagram of Reflektor's core components: the chat client, server, and visualization.

4.1.1.1 Chat Client

The chat client is implemented using the JavaScript open source runtime environment node.js[7]. The application used web sockets to instantiate a chat environment for the users. In addition to operating as a chat client, the application also allows users to *share* and *vote* on the words or phrases from their conversation (see Figure 4.2). While users chat, the server suggests keyword or phrases that it determines to be important to the conversation. These words appear as a list in the Share page (the process by which the server determines which words or phrases are significant is discussed in the Text Parsing and Filtering section below). Users are then able to share or delete the suggested words depending on whether they feel the phrases are important components of their conversation. Once shared, the words appear in both users Vote page. thereby allowing the other participant to vote the phrase up or down. The system is thus a two-step process

[7] https://nodejs.org

of collaboratively determining the content of the visualization and consequently what music is suggested.

Figure 4.2. Reflektor chat client pages.

4.1.1.2 Text Parsing and Filtering

While users chat, each message sent is parsed and analyzed by the server for keywords and phrases. The server is implemented using Django 1.7. and communicates with the chat client using the Django REST Framework[8]. The server processes the messages for each user independently. The server uses text frequency inverse document frequency (TF/IDF) to select keywords from the users' conversation. The server uses the NPS Chat Corpus from the Natural Language Tool Kit[9] (NLTK) as a basis for the TF/IDF algorithm. We determined a threshold for which words and phrases the system chose by processing posts from online music forums. In addition to setting a threshold, we expanded the list of stop words that are included in the NLTK to include words such as album, song, musician, and other words that we found to be consistently present in discussions of music but are not helpful in determining the substance of the conversation.

We sought to enable the participants to communicate about socially significant topics, therefore Reflektor has an addition layer of filtering using social media after the TF/IDF algorithm. If a word or phrase passed the TF/IDF threshold, Reflektor searched the Facebook Search API for corresponding Facebook Pages. If a page is returned with an identical name and the number of

[8] http://www.django-rest-framework.org/
[9] http://www.nltk.org/

“Likes” for the page was greater than 100, the phrase was deemed to be valid. Like the TF/IDF threshold, the Page Like threshold was determined by iteratively processing conversations from online music forums and adjusting the value. This step was crucial because the TF/IDF algorithm can only tell if a word or phrase is common or uncommon relative to a corpus. Filtering the words and phrases through the Facebook API allows Reflektor to determine which of the uncommon phrases were likely to be important to the user rather than simply an uncommon phrase or expression. To illustrate the point, in the sentence: “The new Kanye West album is great” the TF/IDF value for the bigram “Kanye West” is likely equal to the bigram “West album” or “new Kanye”, since each of those bigrams is a phrase that is unlikely to appear in the NPS chat corpus. However, the Facebook API will return a page with numerous Likes for Kanye West but not for the other phrases.

To disambiguate homonyms, the metadata returned from the Facebook API for the keyword or phrase is shown in the users’ Share page. The metadata included the page description and Facebook page category. For some words or phrases, multiple results are returned. In this situation, each result is shown in the Share page and the users can choose the phrase for the category they are discussing. During our calibration of the TF/IDF and page Like threshold, we found some page categories—such as Shopping/retail, Business service, and Pet supplies—returned consistently unhelpful results and chose to omit them.

4.1.2 *Visualization*

After a word or phrase is shared by a user, it then appears in the system’s visualization. We created the visualization using JavaScript. During the study the visualization was displayed at full screen on a 60” wall-mounted monitor (see Figure 4.3). Reflektor visualizes phrases in two different ways: as a mood board or word cloud. The participants saw both visualization techniques during the lab study, but the order was randomly assigned. The content of the visualization updates dynamically as users share and vote on phrases. To select images for the mood board visualization, Reflektor uses the Google Search API. The search included both the phrase and the category from the corresponding Facebook page in the search string. We initially created the mood board visualization using the images form the Facebook Page itself, but while piloting the system, we found the quality of the images to be very inconsistent. After further pilot testing, we felt that using the Google API produced the best mood boards in terms of consistent quality of the images that represented the search phrases.

Figure 4.3. Lab study room set up including example mood board.

To encourage transparency of how the items in the visualization would influence the music being selected, we chose to make the size of the items change relative to the number of votes the phrase receives from users. After being shared, the phrases begin with a default size of the image being approximately 10% of the display. When voted on, the image size doubles to approximately 20% for 2 votes and 40% for three votes. If a phrase has zero votes, it is removed from the visualization. Similarly, the text of the word cloud is constrained to a text box of a size equal to the Each user can vote on the image once and the words and phrases begin with a vote count of one. Thus, the maximum number of votes for a word or phrase is three votes, which would occur if both users vote for the image after it is shared. We chose to allow users to vote for their own suggested phrases so they could remove phrases previously shared or override their chat partner's vote if he or she voted against the phrases they suggested.

4.2 STUDY DESIGN

My collaborators and I conducted a lab study with 10 sets of dyads for a total of 20 participants. During the study, we instructed participants to use the chat client to discuss two hypothetical events and the music they would like to play at each event: a "quiet evening with friends" and a "party". We instructed them to discuss anything they felt was relevant to discuss if they were organizing each of those events. The order for the event they discussed was randomly assigned along with the order for the visualizations they used. Each study session lasted approximately 90 minutes, and included of two rounds of the participants using the chat client to discuss music. The lab session began with the research staff demonstrating Reflektor to the two participants. Between the conversations, the participants rated songs for the event they had discussed and were asked about their experience. After the second round of chatting a longer interview was conducted. We video recorded each study sessions and logged the participants' chat conversation along with phrases they shared, deleted, and voted on.

We recruited participants through online discussion boards and with paper flyers on a university campus. We randomly assigned participants to each dyad, and none of the participants had met prior to the study. Their ages ranged from 18 to 28 with an average age of 21.6. Eight participants identified as female and the remaining 12 identified as male. We compensated participants with a $20 gift card for their time. The study was conducted in a conference room on a university campus with a table in the center and a 60" monitor on the wall (see Figure 4.3).

4.2.1 *Creating and Rating Recommendations*

To evaluate the quality of recommendations produced by the system, we created two sets of five songs for both scenarios the participants discussed. To create the sets of songs we used two techniques: the first technique used a similarity ranking and the second used the system to conducted text parsing on the participants' conversation. In this chapter we will refer to the techniques as *similar* and *text parsing* respectively. To generate the similar songs, prior to the study we emailed each participant a spreadsheet and instructed them to list 10 songs that—to quote from the instructions included in the spreadsheet—"you would want played during a quiet evening with friends" and 10 songs "you would want played at a party with friends". The language was intentionally left broad to encourage the participants to come to their own conclusions about what music would be appropriate for each scenario. We then used their songs to search for music similar to each song in both participants' playlists for each dyad with the Last.fm API[10]. Last.fm's "Track.getSimilar()" function returns the 100 most similar songs as determined from their extensive listening history of its entire user base. We used this function to get a list of the 1000 most similar songs for each participant in both conditions. We then ranked the tracks returned by the number of times they appeared among the songs returned from the other participants' songs.

To generate the songs for the text parsing condition, we searched the Last.fm API for songs tagged with the key words voted on by the participants. The process of creating the recommendations was run once at the end of each of the participants' conversations. We weighted the songs according to the number of participant's votes. When the page for a phrase had the category "musician or band" we queried the last.fm API for the top 10 tags for the musician and incorporated those tags into the song queries. The songs were then ranked and normalized by the number of tags. This allowed us to explore how musicians were used for grounding the conversation, instead of allowing the participants to simply list musicians to generate recommendations.

The lists of similar and text parsed songs were then combined and randomized. We then asked participants to rate the songs on three different rating scales from 1 to 5 based on 1) how much they enjoyed the songs, 2) how appropriate the songs were based on what they had discussed, 3) and how familiar they were with the songs. The songs were presented to participants as one playlist of 10 randomly ordered songs. Participants rated the songs after each conversation and could skip through or replay songs while rating them.

4.2.2 *Chat Log Analysis*

The participants' chat conversations lasted approximately 15 minutes per condition. On average, each conversation had 72 lines of dialog (σ=21.5) and a mean word count of 438 words (σ=110.7). The themes presented in this chapter emerged from the iterative coding and refinement of the participants' chat conversations. My collaborators and I refined and developed these themes by reviewing the videotaped interviews conducted with the participants during each study session. These codes were further developed by drawing on the sociological research presented by DeNora

[10] http://www.last.fm/api

in *Music in Everyday Life* (DeNora, 2000). Between one and three codes were used on each section of the participants' conversation. To ensure that the codes were applied consistently, the authors independently coded the first two interviews and compared the results to determine how consistently the codes were applied. My collaborators and I then iterated and refined codes that were applied inconsistently and reviewed sections of the conversations where variation in codes were found. After revising the codes, the authors then coded one additional interview and compared the consistency with which the codes were applied. After the second round of coding, we achieved an 84% agreement between their application of codes. We then coded the remaining chat logs independently.

4.3 Findings

The findings from the lab study are organized into two sections: the first section discusses our analysis of the participants' chat logs and the themes that emerged in the strategies the participants used to ground their views on music and context. The second section presents the participants' ratings of the recommended music and their views on the associated visualization.

4.3.1 *Grounding Views on Music and Context*

A major goal of this study was to better understand how people determine what music is appropriate and enjoyable for specific contexts. As was mentioned in the methods section, the participants were unacquainted prior to the study. We intentionally recruited unacquainted participants so that they would have to establish what constituted the two conditions of the study—a party or a quiet evening with friends. In this sense the participants were required to establishing common ground by contributing to the discourse of the conversation. In Clark and Schaffer's work on contributing to discourse (Clark & Schaefer, 1989) they suggest that the process of contributing to a conversation has two components: *content specification* and *grounding criterion.* Content specification is the process by which contributors try to specify the content of his or her contribution, and the conversational partners try to register that content. Essentially, one person says something and the other acknowledges it. Once a contribution to the conversation is specified the contributor and partners together try to reach the *grounding criterion*. The grounding criterion is when the contributor and the partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for the current purpose at hand. In this sense, participants iteratively contributed ideas that were established as relevant or irrelevant and then incorporated them into their evolving discussion of what the correct context was.

The participants' process of specifying and attempting to achieve the grounding criterion played out in a variety of ways. In general terms, this process consisted of one the participants specifying what the context should consist of or what music would be appropriate for that context. Then, participants determined if that suggestion was aligned with the evolving notion of the current context. To accomplish this, the participants often utilized one of three criteria: *sonic properties* of the music, *interaction styles* associated with the context of the event, or what *location* enables similar forms of interaction. Each of the criterion is discussed below.

4.3.2 *Sonic Properties of Music and Their Fit for Context*

Discussing what music would be appropriate for the two contexts necessitated that the participants unpack their language to ensure they understood what they were proposing to contribute. This was the case in all the participants' conversations as they tried to determine what constituted a quiet evening with friends, or "a chill party" as it was often described. In the example below, we can see the two participants discuss what the language means and subsequently how the meaning of the words might correspond to what music is appropriate[11].

Bonnie: *cool. so what do you do at chill parties?*
Brian: *Music, drink, conversation...*
Bonnie: *alright. so we probs want music that's easy to talk over*
Brian: *What constitutes chill for you? Slow tempo, repetitive, acoustic...?*

We can see this conversation start off as a discussion of the activities for the party, but then quickly Brian tries to match the word "chill" to sonic properties of the music. This invites Bonnie to provide more details on what chill means to her.

Bonnie: *music that's easy to talk over*
Bonnie: *oh um*
Bonnie: *probably beats that aren't too hard*
Bonnie: *not too many wubs lol*
Bonnie: *you?*

[11] All names in this chapter are anonymous identifiers given to participants during the lab study. The transcripts shown are from the participants' chat logs. Any edits to the chat logs are indicated with brackets.

Bonnie again discusses chill in terms of activities that the music would permit, specifically something that accommodates talking. Brian's does not respond, thereby not validating her contribution, so she further elaborates on her initial response by describing the music in terms of properties of the music. Instead of defining chill, she instead decides to negate things that are not chill. She does this by referencing "beats that aren't too hard" as being inappropriate features of chill music. She then goes on to elaborate that "wubs"—an expression used to describe the bass modulation commonly used in dubstep music—would be uncharacteristic of chill music. The approach that Bonnie applies is characteristic of *generators* in category theory (Lakoff, 1987, p. 15). Bonnie is establishing the idea of what music is part of her category of chill music by defining a rule that can be applied to music to "generate" the members of that category. Her formulation is:

Music without hard beats – "wubs" = chill music

This strategy is obviously a lighthearted attempt to characterize her socially situated construct "chill", but it is specific enough that Brian seems to understand what she means and proceeds to prompt her to continue establishing what music should be played at their chill party.

4.3.3 *Using Activity to Establish Contextually Appropriate Music*

In all the chat conversations, participants discussed what activities they planned to engage in for the given context. This was the most common way that the interaction of the context was framed by the participants. This occurred frequently when participants discussed dancing and what music they felt would be good to encourage dancing. However, we also saw it extend beyond the direct influence that music has to structure physical interaction, such as dancing, to activities that draw on music's ability to establish a tone or ambiance by drawing on the cultural significance of the music. This process is illustrated in the chat dialogue from Interview 4 when the participants discussed the context of a quiet evening with friends. The first few lines of the conversation consisted primarily of the participants making sure they are connected, but once it is clear they are both connected, Dave states:

Dave: *I like jazz and crooners.*

After 8 more lines of the participants discussing games, the following conversation begins in which they try to establish music that corresponds to the game-playing activity that was suggested.

Dan: *jazz huh?*

Dan: *oh id say poker. how could you even play blackjack with friends lol*

Dave: *Poker is fun. Bring some chips and cards.*

Dan: *but lets get a playlist with a lot of jazz piano*

Dan: *yeah [y]eah [w]ill do*

Dave: *Herbie Hancock*

Dan: *do you know if sarah is coming? shes cute.*

Dave: *She might be. I dunno.*

Dan: *how about Thelonious Monk or Art Tatum?*

Dan: *my brother plays sax so always sends me stuff to listen to*

Dave: *Monk might be a bit on the strange side. Same with Tatum. But I would say Tatum or Monk.*

Dave: *What about Miles Davis? not piano but he is pretty good.*

Dave: *Frank Sinatra?*

Dan: *ok well specifics arent super important. we can mix in Sinatra or Ella Fitzgerald too*

Dan: *the important thing is what are we going to be drinking*

Dave: *Whiskey or wine. Both go well with Jazz and Sinatra I feel.*

Dan: *yeah! [I] have a buddy that always drinks old fashions*

From this portion of the participants' conversation, we can see them trying to match styles of music with activities in which they hope to engage. This begins with card playing but quickly evolves into a more holistic context when Dan asks if Sarah is coming to their quiet evening with friends. Because these participants did not know each other, this was evidently an imaginary person and the idea of getting to talk with a person that he thought was "cute" was used as a suggestion for a form of interaction that they hoped to establish—namely flirting. In all, the participants outline card playing, flirting, and drinking as being appropriate modes of interaction. This portion of their conversation illustrates how the activities establish and contribute to the criterion of what is appropriate for the context that extends beyond the sonic properties of the music. Presumably one could play card games to any type of music. But, for the context that they hope to establish where drinking wine, whiskey, and cocktails while people play cards and flirt, the more sophisticated jazz music is viewed by the participants as being appropriate.

One challenge participants encountered arose when the bands one participant suggested were unfamiliar to the other participant. This created a challenge because it became difficult for them to determine how this music matched with the larger context. An example of this came in Interview 5, where participants had clearly different tastes in music.

Eve: *What would you like to listen to for the first part of the party?*

Eve: *maybe something soft? but not something that will put you to sleep*

Eve: *aloe black?*

Ed: *for the relaxing part, I think it would be good to avoid anything too bass or vocal heavy*

Eve: *what would you [recommend]*

Ed: *never heard of aloe black. what sort of music are they? I would recommend something like dj okawari maybe*

Eve: *its definitely not party music, but more listening but good beats*

Eve: *its more vocal*

Eve: *I think it should definitely create a vibe that will be mellow*

Ed: *yeah*

In this example, we can see Eve begin by suggesting a musician that Ed was unfamiliar with and once it is clear that he is unfamiliar then trying to describe it in terms of the sonic properties. Her description seems to work, but this back and forth went on with them each suggesting several bands that the other only seemed vaguely familiar with. In their second discussion, they switched tactics and began their conversation by focusing more on the ambiance and activities of the party.

Ed: *hello*

Eve: *Hey!*

Ed: *for this one, I think we should go for an energetic vibe*

Ed: *though, I'd like to include some activities besides just dancing*

Eve: *Where do you want the party to be hosted at?*

Eve: *I have a big house! We can have different games and maybe food too?*

Ed: *that could be good*

Ed: *you have something like a ping pong, fusball, or air hocky table?*

Ed: *or maybe a dart board*

After discussing the activities of the party, they then began to suggest musicians, essentially reversing their strategy from the first discussion. After their conversation, Eve stated:

Eve, Interview 5: *"I think the strategy was to start talking about the actual event at the beginning and then at end, 'well what kind of music do you want to see'. So it made it a lot clearer in that sense"*

From this we can see that activity is a valuable tool in providing grounding for music. Activity provide enough cultural cues to help establish the styles of interaction that the participants hope to enact. Music is helpful in accomplishing this, but breaks down when one participant is unfamiliar with specific musicians or music styles.

4.3.4 *Location as an Indication of Music Type*

Another strategy that we saw in the participants' approach to establishing common ground was the use of physical locations as indicators of what music could be defined as appropriate for the context. In Interview 9, participants had little overlap in their musical tastes and therefore struggled to determine what chill music would be appropriate. After some discussion of the music, we see Ira instead begin to suggest various clubs as sources of music in order to establish common ground.

Ian: *but there is electronic music that is prett[y] chill too*
Ira: *same era as [F]leetwood [Mac]*
Ian: *[I might] wanna [check] it*
Ian: *the blues [I] mean*
Ira: *yeah: [a local club] specialize s in that stuff*
Ira: *gimme some electronic [recommendations]*
Ian: *yea like [I] went to a few [clubs] [downtown] and they play stuff that is [like] that*

In this example, we can see Ian suggest electronic music, but rather than establish the type of electronic music he would like by suggesting an artist, Ira contributes to the idea of electronic music by associating it with a well-known local club. Ira then asks Ian to make some suggestions to which Ian also provides downtown clubs to establish common ground. This technique seems to be effective in helping them further refine what music would be most appropriate because it causes them to refine the idea of electronic music as including the music genre dubstep.

By using location as a stand-in for a category of chill electronic music, participants use metonymy to reference the music. Metonymy is a principle by which a well-understood or easy-to-perceive aspect of something and use it to stand either for the thing as a whole or for some other aspect or part of it (Lakoff, 1987, p. 77). A common example of this is when the entire movie industry in the United States is referred to as "Hollywood." By referencing the club, Ira can easily allude to the music played there and the ambience. This use of metonymy is sufficient to help Ian further refine the electronic music that would be appropriate.

4.3.5 *Recommendations and Visual Representation*

Now that I have discussed how participants established what music would be appropriate for the different contexts, I will discuss the music that was recommended to participants and the impact of the visualizations on the participants' ratings. As was discussed in the methods section, after the participants chatted for each condition, their conversation was used to generate a playlist of five songs, which was combined with five songs similar to those obtained from the participants prior to the study. In our analysis of their ratings, we used R (R Core Team, 2012) and lme4 (Bates, Mächler, Bolker, & Walker, 2014) to perform a linear mixed effects analysis of the relationship between the visualization type, recommendation technique, and the participants' ratings of the music for *appropriateness*, *familiarity*, and *enjoyment* of the music. As fixed effects, I entered visualization type (mood board or word cloud), recommender type (text parsing or similarity ranking), scenario, and order into the model. As random effects, I included intercepts for participants and for participant dyads. Visual inspection of the residual plots did not reveal any obvious deviation from homoscedasticity or normality for *enjoyment* and *appropriate* dependent variables. The *familiarity* ratings were largely bi-modal with modes at both 1 and 5 ratings. P-values were obtained by likelihood ratio tests of the full model with the effect in question against a model without the effect in question. My analysis found that neither the visualization nor the recommendation type had a significant effect on participants' ratings for any of the dependent measures. A likelihood ratio test for interaction also showed no significant effect on participants' ratings for any dependent measure.

4.3.6 *From Conversation to Recommendation*

Even though we found no difference between the two recommendation techniques, we were somewhat encouraged that after the participants' brief conversation, the system was able to produce recommendations participants rated as equally enjoyable, appropriate, and familiar to music similar to what they had provided prior to the study. However, because of the differences in how these techniques operate, it is worth investigating how the participants' conversations were or were not accommodated by the recommender.

The goal of this project was to explore music recommendations for a given context. We drew on the Facebook API to help with filtering terms in this process, but terms that had Facebook Pages and were therefore socially significant were not always terms that could be easily associated with musical choices. This meant that the strategies that participants found useful in grounding the conversation with the other participant did not always result in improved music recommendations. To illustrate how these strategies influenced what music was recommended, we will return to the participants' conversation and discussion in *Grounding Views on Music*. In Interview 4, we saw participants establish common ground through the use of activities. In this example, we saw the activity discussed—card games—appear in their visualization and ultimately factored into their music recommendations. However, only seven songs on Last.fm are tagged with the phrase "card games" which led to none of these songs being chosen for their playlists. Similarly, in Interview 2, the participants' chat conversation had a total of 409 words. Of these 409 words, 21 words or phrases were eventually shared by the participants—meaning that these 21 phrases appeared in the visualization and were used to produce the music recommendations. Interestingly, despite the Bonnie's vivid description of what chill meant to her, none of those terms appeared in the visualization and therefore were not used to create the playlists. While her phrase "Wubs" did have associated tags on Last.fm, there were no Facebook Pages returned by the initial query so her description went unaccounted for.

The participants from Interview 9 who used clubs to denote which music was appropriate had a total of 22 unique shared phrases from their conversation. Unfortunately, none of the 22 phrases were the names of any of the clubs mentioned to ground the conversation. After inspecting the results of the suggested terms, we found that the clubs mentioned did have associated Facebook Pages but did not have associated tag data in the Last.fm API. Unfortunately, this resulted in a loss of a rich way to explain contextualized music. From each of these strategies, we can see the same result: a mismatch between how context is grounded and a lack of metadata available to support the participant's strategy.

4.3.7 *Visual Representation and Grounding*

My collaborators and I were surprised that the visualization type had no apparent influence on the participants' perceived quality of the recommendations. The system sought to facilitate participants in establishing common ground by visualizing their conversation as it unfolded, thereby making the criteria used in the recommendations more transparent. Transparency in recommender systems has been shown to improve how much users like the results and confident they are in them (Sinha & Swearingen, 2002). By visualizing the keywords from the conversations, we hoped to improve the perceived quality of the recommendations. In six of the sessions, participants expressed sentiment that the mood board provided a useful way to display the content of the conversation.

> **Hope, Interview 8:** *"I think by looking at the pictures you can kinda of see what we were trying to get at."*

The mood board visualization helped contextualize the recommendations and provide some insights into what music would or would not be appropriate to suggest.

> **Dave, Interview 4:** *"If I saw just this and had no knowledge of the conversation I probably wouldn't suggest Beethoven or Miles David, but I might suggest Bruce Springsteen or Tom Petty."*

At times when the other participant was unfamiliar with a musician suggested by the other participant, the mood board helped provide insight into what a musician or band might be like.

> **Eve, Interviews 5:** *"No I hadn't heard of [the band]. [The mood board helped me see who it was. Because we were talking about mellow music, so that could be… That's something we could play at a mellow party"*

However, in several instances, the mood board caused some confusion when it displayed images incorrectly. The most common examples were instances where the image selected by the participant was unrecognizable. One participant compared this to the process of record shopping but only being able to see the album covers.

> **Hanna, Interview 8:** *"It's like if you're buying a record and the record is just the image and you don't know the band so you have no idea what this about."*

Participants were also confused when the system selected a homonym for the phrase that was selected. For example, displaying a house after the participants had discussed the music genre house. The system sought to disambiguate homonyms by using the Facebook category as part of the search string when searching for the images to populate the mood board. However, when the incorrect images were selected and images with meanings that differed from the intended sentiment were displayed, participants felt that the mood board was less useful than the word cloud.

> **Ira, Interview 9:** *"I would say that the photographs are a little more confusing, because of their association with the word. So there are like two levels of association we need to make."*

Based on these comments, and other similar comments, it seems likely that there are simply trade-offs between the visualization types. When the images are clear and familiar to the participant, they are very effective. Even when they are unfamiliar they can in instance be useful for invoking a tone. But, when they incorrectly reflect the intention of the participant, they are more distracting than the word alone.

4.4 DISCUSSION

In this section, I begin with a discussion of why understanding how people discuss music and its relationship to context is important for the design of context-specific music recommendations. Then, based on the strategies used by the participants, we propose design implications for the design of context-aware music recommender systems.

4.4.1 *Why Views on Music in Context Matter*

To begin our discussion of context and music, it is important to note that I am not claiming that all people view music the same way or that there is a "right" way to view music. Furthermore, there is clearly no right way to view how music is related to the context where it is played. However, I do believe that there are ways that people are *likely* to view music and there are ways that are easier or more difficult for computers to respond to the users' views on music and its relationship to the context. Through our exploration of how individuals communicate that relationship, we are able to provide insights into some mechanisms for supporting the interaction that participants engaged in when trying to negotiate what music is appropriate for the context. The findings from this study suggest that when participants create socially situated recommendations, their goal is not to find specific songs per se, but rather to find music that enables or at least does not inhibit modes of interaction. Recommendations that foreground the music over the modes of interaction therefore may not accommodate users' actual goals.

One consequence we encountered from an emphasis toward musicians and away from styles of interaction came when a participant discussed music not known by the other. In every interview, we observed participants suggest bands to ground the conversation. In the instances where the band was unfamiliar to the other participant, this effectively ended the conversation. While this could lead to the design implication that recommenders should incorporate an easier form of previewing music, we instead feel that it reveals a limitation of how recommenders normally operate. When an artist is recommended, the way that the artist is experienced is through their songs. However, the song that is suggested may or may not be representative of their larger body of work and therefore this may create a bias for or against their other music. Instead, we feel it is important social music recommenders provide an overview of how songs fit modes of interaction. In fact, we have seen a trend toward this in how music playlists are organized on music streaming services. Examples of this include playlists Spotify's curated playlists entitled "Gaming", "Focus," and "Sleep". How these methods of organization extend to other modes of social interaction is an underexplored area and offers opportunities for research in recommender systems and interface design that seek to capture the nuanced relationship between the users' views of the music's role in creating context.

4.4.2 *Interfaces for Grounding Styles of Interaction*

Our analysis of the chat logs and the interviews revealed several strategies used by participants in establishing context and what music was appropriate for that context. Our statistical analysis of the participants' ratings found that neither recommendation technique was significantly different and our analysis of the chat logs seemed to indicate a mismatch in how the participants discussed music and how the system was able to accommodate them. Based on these findings, I have developed several considerations for the design of context-specific music recommendation. These considerations and associated challenges are discussed below.

One of the key findings from the lab study was how the participants drew on sonic properties, activity, and location to establish styles of interaction for the contexts. Each of these topics have been explored elsewhere for their role in the production of music recommendations.

However, because of how prevalent these strategies were in the findings, it is worth exploring how recommenders could be designed to emphasize these considerations.

4.4.3 *Presenting Music in the User's Language*

The sonic properties of the music and how they relate to the context were topics that arose repeatedly and vividly in the participants' conversations. While many recommender systems have allowed users to create recommendations based on musical features such as tempo (Baur, Hering, Boring, & Butz, 2011), we did not see participants use this language extensively when discussing music. In fact, only one participant seemed to be interested in engaging in a discussion of the sonic properties of music using the formal language from music theory. The majority of participants drew on more common language to describe the sonic properties, such as Bonnie's use of the phrase "wubs." As was mentioned in the findings, one example that was of particular interest was the use of generators in the descriptions in music. This technique we imagine could allow for a more flexible method of describing the sonic properties of music for users without extensive knowledge of music. This approach has been explored for movie recommendations (Vig, Sen, & Riedl, 2012), and through commercial services such as APM Music[12], but still offers opportunities for research in context-specific music recommendations.

4.4.4 *Support for Social Activities*

Our findings also provided insight into how music and activity are leveraged to create the ambiance of an event. While the notion of defining music recommendations for a given action has been explored in prior work for activities with a distinctly physical dimension, such as dancing or exercising, but how music is leveraged to create the ambiance for other common place events has been less explored. The role that music played in establishing the tone for the participant's events was of interest. In DeNora's work exploring the role of music in everyday life, she refers to "scenes" that are constructed by drawing on music as a cultural material (DeNora, 2000, p. 123). This includes the connotations that the music suggests. Participants clearly hoped to establish specific scenes and used music as a resource in the creation of those scenes. Unfortunately, because metadata does not exist that explicitly links more nuanced forms of social interactions to various

[12] https://www.apmmusic.com

music styles, these connections are unclear at best. A larger issue is that the relationship between the music and its cultural connotations are highly personal, so a generalized dataset of cultural significance would be infeasible to create. However, the extensive body of work in activity detection that has been enabled through personal informatics could be leveraged to create personal significant music recommendations for social interactions.

4.4.5 *Metonymy as a Resources for Playlist Creation*

Participants' use of metonymy to describe music that would be appropriate suggested additional implications for design. While systems such as (Kirk et al., 2016; McCarthy & Anagnost, 1998; Seeburger et al., 2012) have leveraged location as a mechanism for improving music recommendations, their focus has been on the songs played at the location. Instead, our findings suggested that participants hoped to use the location as a resource in meaning making. As was discussed in the findings, the important implication of using the location to ground the conversation was not simply to establish the music, but to metonymically suggest the styles of interaction and ambiance that are afforded by the music at the location. By focusing on the music at the location instead of the larger context suggested, we overlook a rich resource for grounding. The design of future interfaces could emphasize the ambiance of the location over the music at the location as a resource in creating recommendations for groups in social settings.

4.5 LIMITATIONS

An important question that might be asked about our study design is: "Why base a recommender on how individuals discuss music?" We choose this technique because we wanted the focus of the interaction to be based on communication with other people instead of the system. Language is such a rich resource in meaning making that we felt it was appropriate to begin there in seeing how individuals communicated what music was appropriate for a given context. By limiting the interaction of the participants to just computer mediated communication, we had the opportunity to see what is most crucial when participants communicate about context-specific music. Furthermore, by limiting their ability to communicate to just text, we encouraged the participants to unpack their views in a way that may not have been as necessary with a richer mode of interaction. However, because we relied on chat-conversations over a short duration of time, we inevitably limited behavior that would occur in real world setting or over longer periods of time.

Another limitation came from recruiting individuals that did not know each other. The aim again was to require participants to elaborate their views on what music would be appropriate for different contexts; if participants had a personal relationship, the relevance of the music was more likely to be established and therefore not need explicit elaboration. However, by recruiting unfamiliar dyads we likely missed out on more natural forms of interaction that would occur among friends or acquaintances. Despite these limitations, we feel that the findings provide useful insights into the process of creating context-specific music recommendations.

4.6 CONCLUSION

In this Chapter, I presented a lab study exploring how users collaboratively create context-specific music recommendations. This chapter sought to address research goal 4-1: validating a conversational Relating Interface. To validate the system, a study was conducted with 10 dyads of participants using Reflektor, a system my collaborators and I developed that uses a novel technique for producing music recommendations based on users' chat conversations. Our evaluation focused on understanding how individuals discuss context specific music recommendations (research goal 4-1). The evaluation showed that the recommended music was rated as equally appropriate and enjoyable as music suggested using a standard recommendation technique. Our analysis of the participants' conversations while using Reflektor revealed the importance of establishing common ground when determining context-specific music recommendations. To accomplish this, participants drew on several aspects of music including its sonic properties, the locations where that music was played, and the modes of interaction that the music afforded. However, this work also identifies the incompatibilities between the strategies participants used to communicate these aspects of music when establishing common ground and the available metadata required to support them. By identifying these aspects of music and strategies to communicate them as key to creating context-specific music recommendations, the work in this chapter provides useful insights into a previously underexplored area of context-aware recommendation research. Furthermore, it provides evidence of the need for interfaces that enable users to negotiate the relationship between their views on context and the behavior of the system.

Chapter 5. AN EVALUATION OF CONTEXT-SPECIFIC MUSIC RECOMMENDER INTERFACES

In this chapter I build on the findings from the lab study evaluation of Reflektor in Chapter 4. In the Reflektor lab study we explored how individuals discuss context-specific music recommendations. One of the main findings was that how people determine what music is desirable or appropriate for a context is not well supported by current recommender systems. This results from a mismatch between how metadata is structured and how users discuss music. Specifically, we found that people determine what music is appropriate for a given context by explaining the music using common-sense understanding that draws on the styles of interaction they hope to encourage. Unfortunately, the metadata for music does not necessarily capture this common-sense understanding of music and recommender algorithms necessarily rely on this information. Changing the underlying information about music to match the users' strategies may present a prohibitively large challenge, however this does leave open an opportunity to design interfaces so that they better accommodate users' recommendation strategies.

To explore the design of interfaces to better accommodate user strategies in creating context-specific music recommendation, my collaborator and I conducted an online evaluation of interfaces inspired by the findings from our previous lab study. The aspects of music that the participants sought to communicate were the sonic properties, the activities that the music afforded, and the locations where people would experience the music. The strategies the participants employed to communicate the relationship between these aspects of the music and what music was appropriate drew on principles of categorization and cognitive models. This was most clearly the case in how one participant expressed the sonic properties of music through the use of generators.

By drawing on these strategies to design interfaces, we were able to determine how the user perceives the relative strengths of various interfaces for communicating context-specific recommendations. To evaluate these strategies, we designed three interface mockups and had 61 participants evaluate them along with a traditional music recommender interface. The work in this chapter builds on the findings from our previous work and on prior work on recommender interface design that take into consideration context and meaning. By evaluating various interface designs to determine which designs are perceived by users as most effect this chapter contributes concrete insights into recommender interface design.

5.1 Background and Related Work

In the Reflektor lab study, we found that users draw on a variety of strategies to ground their views on music for specific contexts. This included two strategies that are of interest. The first involved refining ideas by adding or subtracting other common terms to describe the relationship between music and context. This practice is akin to the notion of *generators* from category theory (Lakoff, 1987, p. 88). The second strategy relied on the cultural knowledge of other locations or activities to suggest music that corresponded to the styles of interaction that the activities or locations implied. The idea of drawing on a portion of something symbolically to represent it holistically is known as *metonymy* and is commonly used in language (Lakoff, 1987, p. 77). For example, the phrase “the White House” is a common metonym to represent the executive branch of the government, or the government in general. While other examples existed in our findings, the use of metonymy and generator struck my collaborator and me as an important mechanism for conversational ground and novel principles that could be used to improve the design of recommender interfaces.

Inspired by these strategies, we developed several interface mockups that represented them. These strategies drew on principles of category theory (Rosch, 1999) and cognitive models (Lakoff, 1999) and are representative of principles common to communication in general. It is my contention that by relying on these underlying strategies common in communication we can improve the design of recommender systems. Relying on conversational strategies to improve the design of systems has been explored in prior work with intelligent agents. Bickmore and Cassell’s work has investigated techniques to improve social dialog with conversational agents and evaluated how this improves trustworthiness (Bickmore & Cassell, 2001). Additionally, Bickmore’s work with context-aware conversational agents could improve bonding with the system (Bickmore, Mauer, & Brown, 2009). While their work provided valuable insights into how conversational principles can be used to improve the design of agents, our work builds on it by extend examining how these principles are perceived when represented non-anthropomorphically and how it is extended to recommender systems.

The design of recommender systems has been examined from a variety of perspectives. Of particular relevance to this research is Swearingen and Sinha's work on interaction design for recommender systems (Swearingen & Sinha, 2002). Their work helps to highlight the importance of design in influencing the users' perception of the quality of the recommendations. Work on music Tag Genome project illuminated the value of using tags to improve recommendations through the knowledge they encoded (Vig et al., 2012). Our work builds on this prior research of recommender systems by contributing an evaluation of how user perceive the utility of a variety or communication principles and their influence in creating recommendations.

5.2 Interface Design and Evaluations

Based on the insights from the work in the Chapter 4, brainstormed numerous designs and evaluated them informally for how well they fit the methods we hoped to emulate and how reasonably they worked as a recommender interface. After designing and evaluating several concepts, we choose three interface designs that we felt best enabled users to draw on the strategies of interest. The first design was based on the generator principle, which allowed users to add or subtract different tags to create playlists. We struggled with a way to incorporate metonymy into an interface design, because it explicitly draws on a mutual understanding of the items being referenced and in an online study we felt this could cause confusion if participants did not share the same understanding of music as we intended. Ultimately, we created the *Scenario* interface, which allows users to create scenarios as a mechanism to express meaning around the style of interaction they hoped to instantiate through the music. Finally, we developed the *Graded* interface, which allows users to suggest tags on a gradation of relevance, rather than all or nothing like the Generator interface, or on the activity like the Scenario interface (see Figure 5.1).

To evaluate the interfaces, we conducted an online survey with 61 participants recruited via Amazon's Mechanical Turk. In the survey, the participants were shown approximately 30-second long narrated and subtitled videos for each interface. The order of the videos was randomly assigned. After watching each video, the participants were asked to rate the interface on a Likert scale for: ease of use, ability to make recommendations specific to their music tastes, appropriate for different event or activities, and allowed the users to express what they think about music. These questions allowed us to assess how the interface performed relative to the key principles of expressing ideas in the user's language, enabling forms of interaction, as well as accomplishing basic recommender functions of being easy to use and providing useful recommendations.

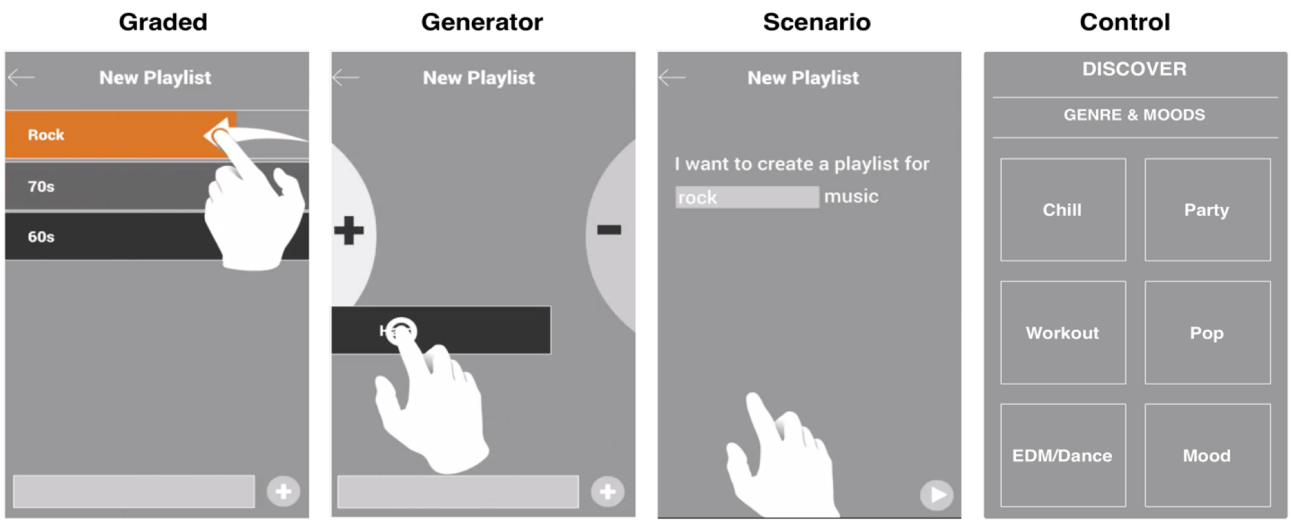


Figure 5.1. The three interfaces evaluated and the control interface.

We paid participants $1.50 for completing the survey, which took approximately 10-15 minutes. We used Mechanical Turk's internal filters to limit participants to only those from the United States with a prior task completion approval rating greater than 95%. The survey was designed so that participants could not proceed without completing all required questions. Additionally, we asked participants to state how often they listened to 20 music genres from "daily to weekly" to "less than annually". We also included an option to state "I don't know this genre". In these 20 genres, we included some of the most popular genres per Nielsen's 2015 U.S. Music Report (Crawford, 2015). We also created two fake genres (romp and fighthall) that were words or phrases not present among the 944 genres listed by Spotify[13] and that returned no results from the last.fm API[14]. We initially used a third fake genre, glamcore, but after reviewing results we determined that it was too close to the genre "glamrock" and decided to include participants that stated they listened to this genre. Participants that stated they listened to the other two fake genres were deemed to be either too unfamiliar with music or not actively participating in the study and their survey results were omitted from the analysis. After filtering the participants, we had a total of 61 valid responses. The demographics of the participants were diverse, but were skewed slightly

[13] https://news.spotify.com/us/2009/03/24/spotify-genres-the-full-listing/
[14] http://www.last.fm/api

toward younger, more educated males. In total, 39 participants identified as Male and 22 as Female. For education, 31 reported having a college or graduate degree, 25 had some college education, and five reported having a high school education.

5.3 SURVEY FINDINGS

The survey asked participants to rate each interface for the four criteria detailed above: *ease, specificity, appropriateness and expressiveness* (see Figure 5.2). We chose these criteria because we felt that they characterized the design goals of creating personal and contextual recommendations. I conducted a non-parametric Friedman ranked sum test of differences on the ratings and found a significant effect for ease (Chi-square=6, p<.05), specificity (Chi-square=10.63, p<.01) and appropriateness (Chi-square=15.5, p<0.001). A Wilcoxon ranked pairs test revealed that the Graded interface was significantly easier to use than the Generator interface (Z=436, p<.05). The Scenario interface was found to be significantly less specific than the Graded (Z=514, p<.05) and Generator interface (Z=515, p<.01). Interestingly, despite being seen as easier to use than the Generator interface and significantly better at producing specific playlists, the Graded interface was rated as significantly *less* appropriate than the Scenario interface (Z=142,

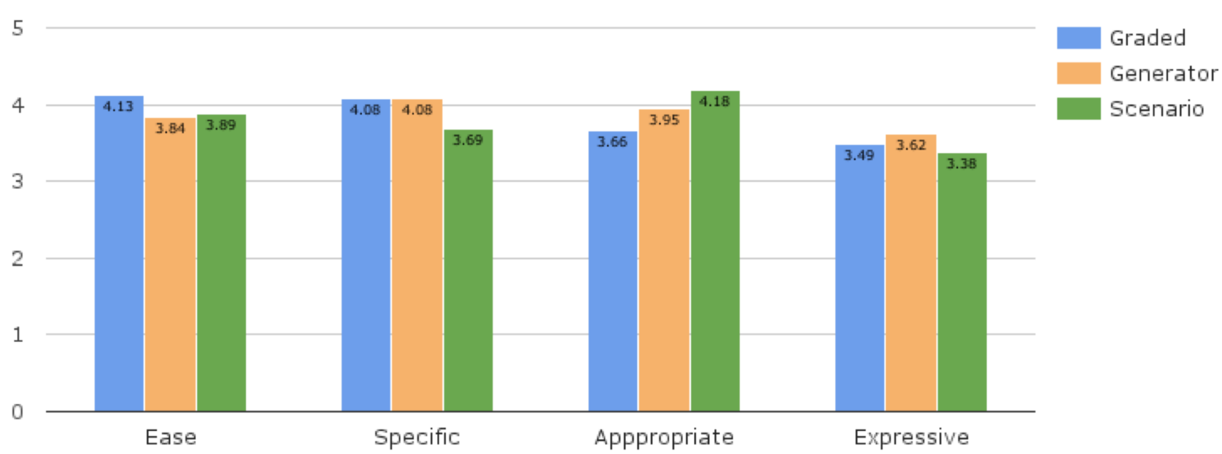


Figure 5.2. Average interface ratings.

p<.005) and the Generator interface (Z=103.5, p<.05).

The participants were also asked to select the interface they felt was best for each of the four features *ease*, *specificity*, *appropriateness* and *expressiveness.* In addition to the Graded, Generator and Scenario interface we included a "traditional" recommender interface. The design of the traditional interface was based on a wireframe of the current interface used in the Spotify mobile application (see Figure 5.1). The results showed the that the traditional interface was widely considered the easiest to use (see Figure 5.3). However, for specificity, expressiveness, and overall preference the participants choose the Graded interface most frequently. Interestingly, the Scenario interface was chosen far more regularly for the appropriateness (see Figure 5.3)

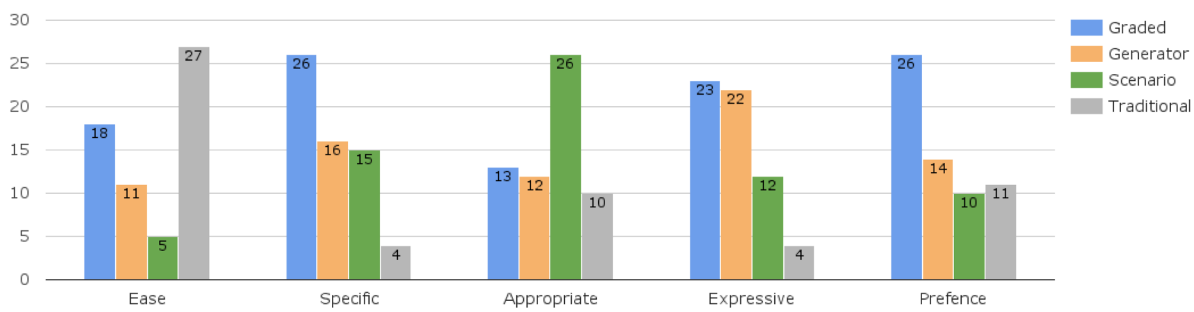


Figure 5.3: Interfaces selections.

5.3.1 *Participants' Comments on Interfaces*

In addition to the participants' ratings of the interfaces, they provided comments on their preferred interface. After reviewing their comments for themes, I applied up to two of the themes as codes on each comment. Two codes were used, because the comments were generally short (word count $\mu=24.6$, $\sigma=16.4$) and the sentiments expressed only reflected one or two of the codes developed. Of the themes present in the comments, the most common overall was ease of use, which appeared in 22 of the 61 comments. This theme was the most common amount among participants who preferred the traditional interface; it represented the explanation for 8 of the 11 participants (73%,) who preferred the traditional interface. Ease of use was the second most common reason for the Graded interface (N=7), but this only represented about 27% of the comments made. Ease of use was the stated reason approximately 30% of the time for Scenario (N=3) and Generator (N=4) interfaces. This was somewhat surprising, because the Graded interface was the preferred interface by a wide margin, but was not viewed as easiest to use, which was the most common theme the participants discussed when they explained their decision. When commenting on the ease of use of the Graded interface a participant stated:

Participant 60: *"The one with the tag adjustment is the most unique - and I think it would be the most effective at making playlists I'd want without sacrificing simplicity or spending too much time. When I create a playlist - I already want to be listening to music - so I'm not going to use anything that takes a long time - like the add and remove tags - for example."*

Based on this and similar comment it seems likely that the Graded interface was viewed as being easy enough to use, but they excelled in other ways that made it the preferred interface.

Customizability and personalization was a close second for most prevalent theme in the comments (N=21). This theme was the most common explanation for participants that preferred the Graded interface. Approximately 69% (N=18) of the comments for participants who preferred the Graded interface listed customizability and personalization as the reason for their selection. While customizability and personalization are commonly cited as important heuristics in interface design and recommender system design, the extent to which this influenced the participants' choice was surprising. When commenting on the ability to customize recommendations, one participant stated:

Participant 40: *"The [graded] interface is right in the middle of what I feel would provide the best listening experience. It is not too broad or too specific. It provides the user the ability to refine a music playlist without being too restrictive."*

The findings seem to suggest that the participants valued ease of use and personalization, almost equally, but that the graded interface was not viewed to be the easiest interface to use. However, we can see that the tradeoff between ease of use and personalization was worthwhile thereby making the Graded interface the preferred interface. One participant summed this sentiment up by stating:

Participant 21: *"To me it just seems like the best. Perhaps not the simplest option - but I personally love giving weight to each category and adding or subtracting percentages of what I'd like from each tag/genre/subgenre."*

The only other theme that appeared more than a handful of times in the comments was the perceived precision of the interface in making recommendations. This theme appeared in 17 of the comments and participants that cited precision as their explanation for choosing an interface were largely drawn to the Generator interface. In all, about 45% of the participants that cited precision as a factor in their choice selected the Generator interface. This suggests that participants viewed this mechanism as being their preference for detailed refinement of music playlists. An example of this mentality came when a participant stated:

> ***Participant 28:*** *"I like the ability to make known what I dislike. a lot of genres overlap or are similar - but I might want to exclude triphop but keep hiphop - for example. In the other interfaces I think I would be stuck with both."*

5.4 Discussion and Conclusion

The participants' choices and ratings of the interfaces suggest that while ease of use, specificity, appropriateness, and expressiveness matter, the clear preference is for interfaces that enable the user to customize and personalize their recommendations while being easy to use. In this sense, the clear preferred interface was the Graded interface. This seems largely consistent with the ratings that found the Graded interface to be significantly easier and more specific to use than the Scenario interface, as wall as easier to use than the Generator interface. The fact that the Generator interface was viewed as being significantly less appropriate for making recommendations than the Scenario and Generator interface is somewhat surprising. The Generator interface was still the clear preference; this seems to suggest that appropriateness was not a key metric used by the participants when evaluating the interfaces. However, specificity was an important factor to the participants and this factor seemed to be at odds with ease of use. The tradeoff between the two seems clear: the more adjustments that an interface affords, the harder it is to use but the more refined the outcome should be. Given this tradeoff, it seems reasonable to assume that in the design of future interfaces the balance between specificity and ease of use should be optional to the user so they are given the freedom to adjust when preferred or avoid when not preferred.

There was a clear relationship between the comments provided by the participants and the evaluation criteria we asked them to use. Most of the themes that emerged are at least similar to the evaluation criteria used for the recommender interfaces in the study. It is unsurprising that after being asked to evaluate interfaces in a specific way the participants would couch their comments in these terms, but it surprising that this was not the case for personalization. While personalization was not a component that we had intended to explore, based on the users' preferences it seems that this is major factor in how users evaluate music recommender interfaces. Personalization is a well-known factor in the evaluation of recommender (Setten, Instituut, Mcnee, & Konstan, 2005) systems, but we were surprised to see the role that it played in the participants evaluation here. Given this, it seems clear that in the design of future interfaces, the ability to personalize results should be paramount.

In this chapter, we conducted an online evaluation of three interfaces that each represented a common metaphor for cognitive models. This chapter provides design guidelines for interfaces to communicate context-specific music recommendations (research goal 5-1). Our evaluation suggested that drawing on Graded categories provides a preferred balance to users in terms of personalization and ease of use. We also found that interfaces that draw on models of interaction result in interfaces that are viewed as being able to produce significantly more appropriate music recommendations. This chapter contributes insights into how users perceive principles of cognitive models influence on recommender systems as well as the tradeoffs between these principles. These findings provide support for the hypothesis that recommendations from Relating Interfaces will better reflect user perspectives than recommendations from non-relating interfaces. However, the findings from this chapter only represent participants' expectations; in Chapter 6 we will evaluate this hypothesis in a field evaluation to determine how well their expectations match their experience using a Relating Interface.

Chapter 6. FIELD DEPLOYMENT OF A CONTEXT-AWARE MUSIC RECOMMENDER INTERFACE

Technology has enabled a proliferation in the availability of music such that there are nearly unlimited opportunities to find the right song for a given context. Context is a rich resource to draw on for meaning making and a great deal of work in HCI has attempted to improve recommender systems by leveraging aspects of the users' context. While drawing on aspects of the users' context clearly creates opportunities to improve recommender systems, how users' feel about their context and how this influences what is appropriate for that context is a key challenge for making the right match. As technology becomes increasingly able to detect and respond to the user's context, allowing users to negotiate the relationship between their context and how systems respond is a topic of growing importance. [15]

To investigate this topic, my collaborators and I conducted a three-week field deployment with twelve participants to evaluate a music recommender that we designed and developed; we call the system *Harmonizer*. Harmonizer is a context-aware music recommender system with an interface, which let users adjust the weighting of songs used to create music recommendations for their context. To evaluate the system, we compared participants' ratings of playlists recommended by the system when using the interface and when using a control interface. We also compared the participants' rating of playlists created using the weighting from the interface to a control set of similar songs for their context. This allowed us to test the following hypotheses: H1) Using the interface will improve the participants' ratings of the playlists they create relative to recommendations that only use their context, H2) The weighting from the interface will improve the participants' ratings for the playlists relative to recommendations that only use their context.

This work builds off the findings from the lab study in Chapter 4 by providing an evaluation of a context-aware music recommender designed based on the strategies surfaced by that work. It further extends the interface evaluation conducted in Chapter 5 by exploring how these interfaces perform in naturalistic settings. This work contributes a new technique for providing music recommendations, which participants rated as more enjoyable than context-aware music recommendations or similar songs alone. Furthermore, this work contributes insights into how

[15] The work in this chapter was done in collaboration with Julie Kientz and Aubury Jellenek. Jellenek was primarily responsible for interface design and development on the Harmonizer application.

users negotiate the relationship between their context and the music they prefer for their context. This provides evidence of the utility of Relating Interfaces to allow users to communicate how music recommendations should be personalized to their context.

6.1 Background and Related Work

Leveraging context offers unique opportunities in how systems respond to users. However, challenges have been noted in how context is interpreted and how this matches the lived experience of the context systems detect (Salvador & Anderson, 2003). This mismatch has led some to call for enabling users rather than calming them (Rogers, 2006). Enabling users suggests offering an interface for interaction with the system rather than relying agency of the system alone. Direct manipulation interfaces have been lauded for their ability to reduce the distance between users' intention and the system's behavior relative to command line interfaces (Shneiderman, 1983). One key limitation arises from their attempt to amplify users knowledge of the domain by encouraging them to think in terms of the application domain thereby requiring a finite vocabulary that is clear to the user (Hutchins & Hollan, 1985). While this can be helpful when the scope of the application is well defined, context-aware systems are difficult to scope and the lived experience of the context may differ from the systems' ability to detect and interpret it.

In this work we hope to address this limitation by evaluating an interface that aims to allow users to negotiate the relationship between their context and how the system responds. For this evaluation, music is a particularly appropriate phenomenon to evaluate relative to context because of the unique role music plays in shaping interaction and behavior (DeNora, 2005). A variety of context-aware music recommender systems have been designed and evaluated. One important component of context in relation to music is the users' mood. Using mood, or affect, to curate playlists is a topic that has been explored by analyzing users galvanic skin response (Healey, Picard, & Dabek, 1998) or through speech analysis (Bauer et al., 2011). Both of these systems however, relied on proactively detecting the users' mood and did not allow the user to update how the system responded. Other systems have used manual selection of mood to retrieve music, through the use of musical signal process and the Ontological Web Language, a web language that allows authors to develop a formal way to describe taxonomies and classification networks to structure knowledge from various domains (Han, Rho, Jun, & Hwang, 2010). This technique is intriguing, but again is limited by pre-existing relationships that the user has no ability to influence.

Our work contributes to these systems by providing a mechanism by which users can influence how the system responds to the affect detected or provided by the user.

Other aspects of the users' context have been drawn on to create recommendations including their location, (Kukka & Patino, 2009) or the presence of others (Baumann et al., 2007; Crossen et al., 2002; McCarthy & Anagnost, 1998). While this prior work provided important insights into how users experience music and their context, this current study contributes to it by focusing on the interfaces that users can use to negotiate this relationship. Furthermore, this work provides an empirical evaluation of the playlists created by the system, which adds validity to the strength of the findings.

6.2 System Design

Harmonizer is a mobile web application and recommender system that was designed to provide users with context-specific music recommendations. To accomplish this, it must first determine what music the participants listened to for various contexts. Then the system must detect and provide music recommendations for those contexts. To determine what music the participants preferred for different contexts, the system logs the users' music listening history and sends them surveys to complete. The surveys include questions about their context while they listened the music. It then uses this information to create recommendations for those given contexts, which the user can access through the recommender interface. The web interface and the server the system uses to accomplish this are discussed in detail below (see Figure 6.1).

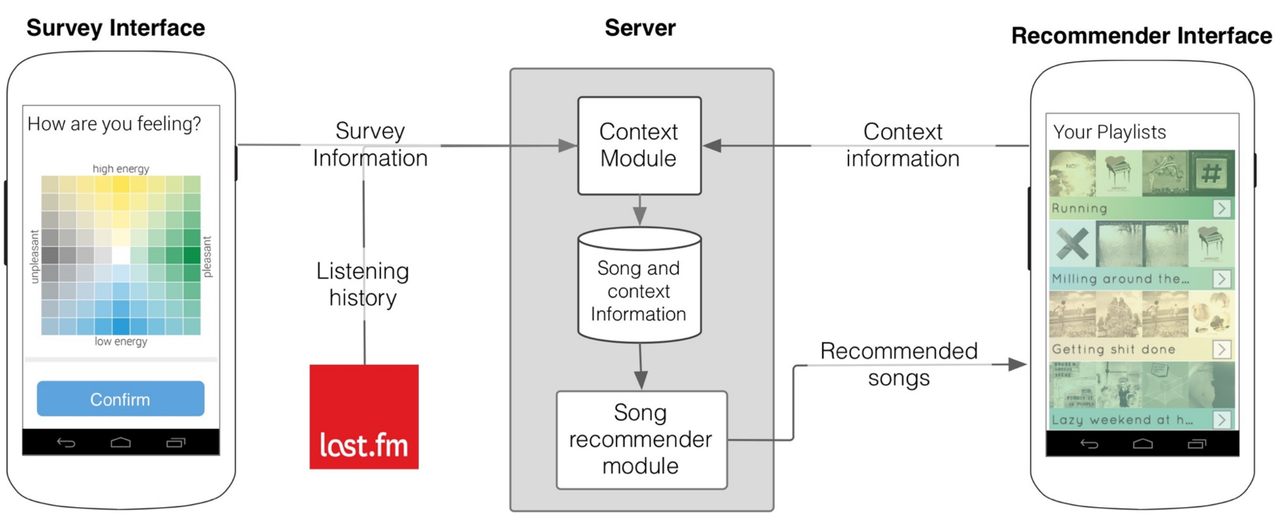


Figure 6.1. System diagram for Harmonizer.

6.2.1.1 Survey Interface and Context Module

The server is responsible to determining when users listen to music and then sends them surveys to complete. During the study, the server checked the participants' music listening history every two hours from 9am to 9pm. The participants' music listening history was recorded on the music logging service Last.fm. If the system detected that the participant had listened to more than four songs consecutively, it would generate a survey. We choose to look for multiple consecutive songs to ensure that the participant was likely to be engaged in the same type of activity while they listened to all the music in each survey. The number of songs chosen and the duration of time between songs that were deemed consecutive were based on informal testing and the work of (Ragno, Burges, & Herley, 2005) on expert authored music streams. If more than 12 songs were played in a row, the songs were broken up into multiple surveys. If a survey was generated for a participant, they were sent a text message that contained a link to any surveys they had not completed. This meant that participants were sent a maximum of one text message every two hours regardless of the number of surveys they had to complete. To ensure that the participants could accurately recall their information for the survey, at the end of each day any incomplete surveys were deleted.

The survey had four main pages each of which prompted the participant to provide information about their context (see Figure 6.2). Each of these pages were required and appeared in the same order. This included the *Label Page,* where the participants were asked to provide a label for what they were doing when the listened to the music logged by the system. No labels were included, and the participants were asked to "provide a label that best describes what you were doing when you listened to the songs below". The participants could manually enter labels, but after they created a label it was added to a list of previously used labels which they could select when completing surveys. The *Editor Page* was where they could remove any songs they did not want included by their recommendations. In the *Map Page,* participants submitted their location using the Google Map API. Their location was recorded as GPS coordinates. The final page of the survey was the *Mood Page,* which allowed participants to enter their mood according to an adapted version of the affect grid (Russell, Weiss, & Mendelsohn, 1989). Their mood was recorded as X, Y coordinates in whole integers.

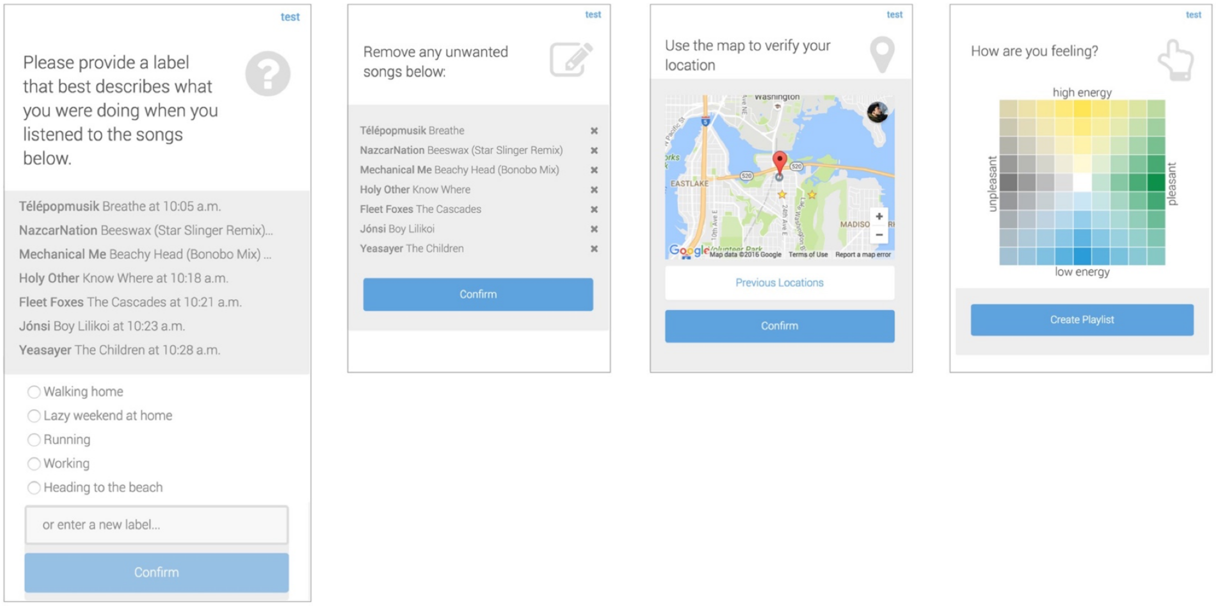


Figure 6.2. Interfaces for the Survey portion of the application.

To provide context-aware music recommendations for the participants, we analyzed their survey responses and the associated listening history to create a collection of similar songs to draw from and a context suggesting when those songs would be appropriate to play. For the purpose of clarity, we will refer to the collection of similar songs and the associated context as *Context-Playlists*. One Context-Playlist was created for each label participants provided that had at least four associated survey responses. The survey data was averaged to create a profile for that label, which was then associated with similar music. To average the survey data, we used K-means clustering to find the top two centroids for each label and used these as indicators of the top two moods the participant was likely to be in when listening to that music. Similarly, we found the top centroids for the participants' GPS data and used this location as indicators of places where they were likely to listen to the associated music. This information was then stored as a vector with three dimensions: two moods and a location for each label. To associate the music with each label, we used the Last.fm API to find the top 60 similar songs to the music the participants listened to most for each label.

In addition to collecting the 60 similar songs for each Context-Playlist we also found the top six songs the participant had listened to for that label while completing the surveys. We called these six songs the *seed* songs, because they were used in the interface to filter for the most similar songs in the pool of songs for each Context-Playlist. To accomplish this filtering, we gathered the top 30 tags applied to the similar songs for each Context-Playlist from the Last.fm API. Instead of simply choosing the most common tags on the similar songs, we instead normalized the tags by how commonly they were applied to any songs and by the number of people that had used that tag according to the Last.fm database. Doing this allowed us to reduce the influence of the most common tags (such as Rock, Hip Hop or EDM) and instead find the tags that were most common and most unique to the similar songs in each Context-Playlist. This approach to finding the most relevant tags for this collection of songs drew on the work of (Sen, Harper, LaPitz, & Riedl, 2007). Once the tags were determined for the similar songs, the normalized number of times each tag was applied to each song in the Context-Playlists was added to all the similar songs as a value between 0 and 1. This information about the distribution of tags on each song was then used as the basis for selecting songs when playlists were made by the participants.

6.2.1.2 Recommender Interface and Recommendation Module

Once the Similar-Playlists were created, the participants could access them through the web application (see Figure 6.3). The recommender asked them to first submit their location and mood using the same method as the survey. The system then used cosine similarity on their submitted location and mood to predict find the most similar Context-Playlist to their current context. The participants were then shown lists of their Context-Playlists and are free to select whichever playlist they prefer, but their choice was logged for comparison later. Each of the Context-Playlists are represented by the top album covers based on their listening history when they completed the surveys. The album covers were also desaturated and recolored using the colors from the top two moods reported for the Context-Playlist. In this way, we hoped to inform the users about the music and the associated affect from that Context-Playlist.

After selecting a Context-Playlist, the participants are then shown the Harmonizer, which displays the seed songs as sliders. Each of these songs act as a weighting mechanism, which the users can adjust by moving the slider further to the right to signify a greater weighting or to the left to reduce the weighting. This weighting is then summed and averaged for each tag on the seed songs. This is then used as a vector and compared to each vector of song tag data in similar songs for the Context-Playlist. Therefore, by adjusting the seed song weighting in the Harmonizer the participant is able adjust the socially aggregated data and then find the most similar songs to it. The selected songs are then embedded in a Spotify playlist in the final page; from this page the participants can rate the songs. In addition to embedding the playlist locally, the playlists were added to the participant's Spotify Playlists for later access.

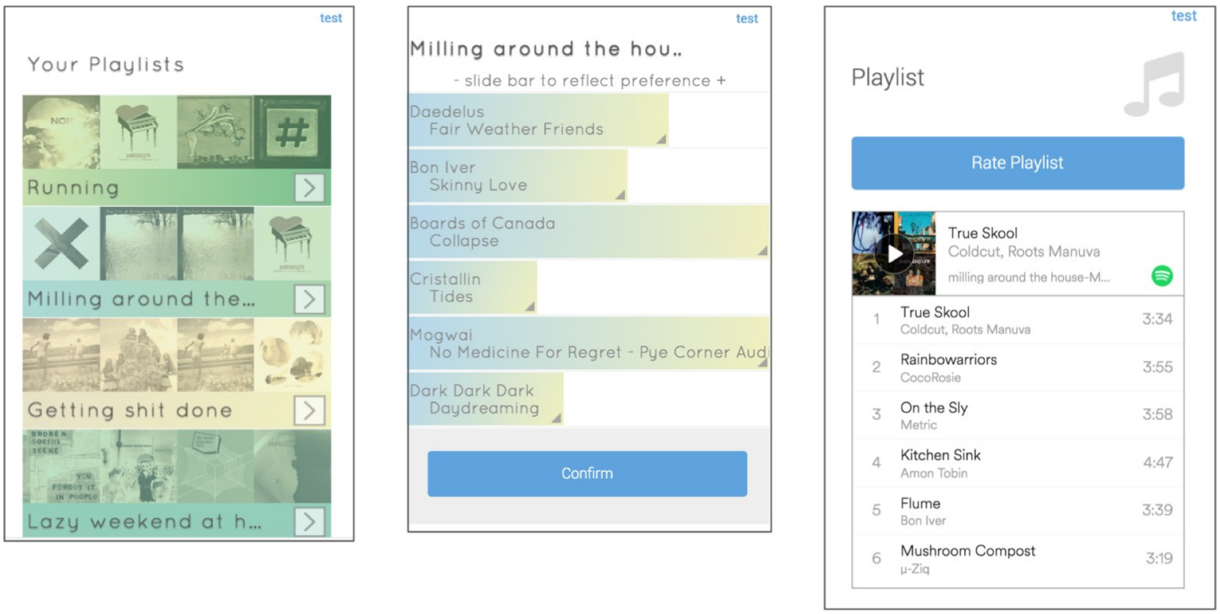


Figure 6.3. Recommender interfaces. Left, the participant's playlists. Center, the Harmonizer. Right, the generated playlists.

6.3 FIELD DEPLOYMENT

To evaluate the system, my collaborators and I conducted a three-week long study with 12 participants. During the first two weeks of the study, the participants were asked to listen to music normally and the complete surveys. In the final week of the study, the participants were asked to use the recommender system to create and then rate recommended playlists. The participants were recruited from the online message board Reddit[16]. We screened the participants by asking how often they listened to music and rejected participants that, during their screening, did not state that they listened to music several times daily. All the participants were university students and their ages ranged from 18-29 years (μ=20.33). Nine participant identified as male and the remaining three identified as female. The participants stated that, on average, they listened to between two and six hours of music each day (μ=3.6) The participants were provided with a month of free access to Spotify music. Additionally, each participant had a Last.fm account created for them

[16] https://www.reddit.com/

which was set up to log all the music played in their Spotify account mobile and on their personal computer.

Prior to the deployment we conducted an initial in-lab study session with each participant. During the session, I walked them through the study design and how the system worked. When necessary, I also created a profile to log their music on Last.fm. If participants did not have one, I also created a Spotify account so that the participant could listen to music freely during the study. I then tested sending them text messages and logging their music to ensure they could access and complete surveys on their phone. I walked them through the completion of one survey and provided an explanation of how to report their mood on the affect grid. They also completed a questionnaire of basic demographic information during this session. This meeting generally took less than 60 minutes.

For the following two weeks, the participants were asked to listen to music normally and complete any surveys they were sent. They were sent surveys up to every 2 hours from 9am to 9pm. As was mentioned in the system description, the server parsed their music listening history in Last.fm and sent them a survey if they had listened to more than four songs consecutively. If a survey was generated the participants were sent a text message with a link to the survey. Participants could complete the surveys at their leisure, but at the end of each day any incomplete surveys were removed. The surveys, on average, took less than 1 minute to complete. We asked participants to complete at least 4 surveys per-day with the hope that we could collect 40 or more surveys over the course of the two weeks. In some cases, at the end of the two-weeks participants had not completed enough surveys and were asked to continue for an addition few days to complete enough surveys so that we could make recommendations for them. When this occurred, we compensated the participants with an additional month of Spotify music.

Once the participants had completed enough surveys to create music recommendations, I processed their survey results and their related listening history. This information was then organized as a Context-Playlist. After the participants' survey information was processed, the participants were asked to use the recommender a few times daily to create playlists and then rate the playlists for enjoyment and appropriateness. As was mentioned in the system description, the participants were asked to report their mood and location. This information was used by the system to predict a Context-Playlist for the participant. The prediction was based on the similarity to their reported mood and location from their surveys. The participants were free to select any playlist, but we recorded our prediction and their selected Context-Playlist. After they selected a Context-Playlist, they were randomly assigned to either see the Harmonizer interface or a control, that just displayed the seed songs used to create the recommendations. If they received the Harmonizer interface, they could adjust the weight of the seed songs and then create a set of recommendations. In the Harmonizer condition, the participants' weighting information from the seed songs was randomly assigned to be used or instead songs were selected at random from the similar songs for that Context-Playlist. The playlist of songs was always 6 songs long and displayed in a Spotify playlist, which they were asked to rate and comment on.

After the week, I conducted an exit interview with each participant to discuss their experiences in the study. Seven of the interviews were conducted in person in an office on the campus of University of Washington and five were conducted via teleconference. The in-person interviews were video recorded, and the teleconference interviews were recorded via screen capture or audio alone depending on the preference of the participants. The interviews lasted approximately 60-minutes and at the conclusion the participants were paid $50 dollars provided via PayPal for their participation in the study.

6.4 Findings

The findings are organized into two sections: the first section discusses the analysis of the participants' playlist ratings. The second is an overview of the main themes that emerged from the interviews with the participants, their views on the system, and how it impacted their ability to discover and curate music. These sections are discussed separately below.

6.4.1 *Playlist Ratings*

To analyze the participants' playlist ratings, I used R (R Core Team, 2012) and lme4 (Bates et al., 2014) to perform a linear mixed effect analysis of the relationship between the interface condition (Harmonizer or list interface), recommendation technique (Weighting or random), the context (Predicted or not predicted), and the participants' ratings of the music for *appropriateness* and *enjoyment*. As fixed effects, I entered interface, recommender and context type into the model. As random effects, I included an intercept for participants. Visual inspection of the residual plots did not reveal any obvious deviation from homoscedasticity or normality for the *enjoyment* and *appropriate* dependent variables. P-values were obtained by likelihood ratio tests of the full model with the effect in question against a model without the effect in question. Users ratings for enjoyment were found to be significant for interface type (Chi-square=5.5874, P=0.018), but not context or recommendation type. No significant results were found for the appropriate measure on any of the independent variables. Also, no interaction was found between the independent variables. This indicates interface type significantly influenced the participants' rating for their enjoyment of the playlists the system produced, but not their appropriateness. Therefore, these findings suggest rejecting the null hypothesis for H1: Using the interface will improve the participants' ratings of the playlists they create relative to recommendations that only use their context, but not for H2: The weighting from the interface will improve the participants' ratings for the playlists relative to recommendations that only use their context.

6.4.2 *Interview Findings*

I conducted an exploratory analysis of the interviews by reviewing each interview looking for participants' comments that would illuminate the findings from the quantitative analysis. This included insights into their experience using the interface to create playlists, the differences between appropriateness and enjoyment, the balance between refinement and randomness in the recommendations, and the relationship between their context and their music choices. These topics are discussed in detail below.

6.4.2.1 Creating Recommendations with the Interface

One of the main findings from the quantitative analysis was that participants rated the playlists created with the Harmonizer interface as being more enjoyable than the control interface. The analysis of the interviews shed some light on why participants felt that the results were more enjoyable. Even though the actual playlists were not always influenced by their use of the interface, participants felt that the interface gave them an improved ability to refine the recommendations. Participant 7 noted that:

> **Participant 7: "***I think the most recent ones I was doing was [playlist name], and that had a Japanese and Danish bands and I felt like listening to something similar but you definitely nailed it with how it picked songs. I already discovered new music and now bands. So I was blown away with how well it worked."*

From his comment we can see that the Harmonizer accomplishes two things. First it gave the participant the ability to customize the playlist, and secondly it does this using language the participant is familiar with. They can see the songs as proxies in a way that the system could respond to when providing the recommendations. Thereby allowing them to include or exclude songs according to their adjustment of the seed songs.

Another important component of their enjoyment of the recommended playlists stemmed from their ability to refine the playlists to the extent they felt was appropriate. Some participants stated that they used the sliders as binary measures to completely exclude items:

Participant 1: *"Yeah there were definitely times where I would slide them off and I would get what I wanted."*

While others took a more graded approach when adjusting the weighting.

Participant 11: *"It was kind of satisfying to be able to smoothly scroll… it seemed very fluid. There weren't check marks saying "do you want this song or not". It was on this scale, so I like this aspect of it."*

From these quotes, we can see that the interface was found to be useful for either strategy without being unwieldy to the opposing strategy.

Participants had mixed opinions on the use of songs as the weighting element in the interface. While some felt that it was clear and provided multiple elements of weighting simultaneously, others felt that the granularity of using songs was not specific enough and using genre would have been their preference. This sentiment was expressed by Participant 1 who stated:

Participant 1: "*Cause that's what I was doing, I was being like "well I don't feel like listening to k-pop I'm gonna take it all off, or I'm just going to leave the hip hop on." And make a hip-hop playlist right now… with a genre slider it would have been much much easier."*

Conversely, Participant 12 felt the opposite and felt that artist would have been a better filtering mechanism. He stated:

Participant 12: *"Yeah for me personally I think it was a little too granular to use the songs*."

Given these differences, using the songs as the weighting object provided at least enough detail most of the time, but clearly more work could be done in this area.

6.4.2.2 The Difference Between Appropriate and Enjoy

One of the more interesting findings from the quantitative analysis was that while the participants rated the playlists they created with the interface as more enjoyable, but they did not rate them as significantly more appropriate. During the interviews, we asked the participants to unpack their views on the difference between music that is appropriate versus enjoyable. One common comment was that the recommendations were not wrong for their stated mood, but that they didn't coincide with the mood the participant was trying to achieve. Participant 7 summarized this sentiment by stating:

> **Participant 7:** *"For the most part it fit the mood properly. It was just that there was an error in establishing what mood am I currently in versus what mood am I trying to get to."*

Several participants mentioned this issue may have stemmed from how they reported their mood. During the survey portion of the study they were encouraged to report their mood when they listened to the music, not report the mood they were trying to achieve by listening to the music. Participant 6 expressed the difference this way:

> **Participant 6:** *"Mum, I felt like at the time it was ok, but now I'm seeing the survey could have used not just what your mood was, but what you wanted your mood to be."*

Clearly the participants felt that there was a discrepancy between the mood of the music and the mood that they hoped to achieve. Based on these comments it seems this was a factor influencing how appropriate their music choices were.

6.4.2.3 Refinement and Randomness

Our analysis of the ratings did not find a relationship between the participant ratings and the recommendation type used to produce the playlists. Although whether the recommendation used the weighting was random, the songs were always chosen from the pool of similar songs for the participant's selected Context-Playlist. As a result, this injected some randomness into the recommendation process. While the participants seemed to generally appreciate the interfaces' ability to produce specific recommendations, at times they also seemed to appreciate the randomness as well. This contributed to the interface being useful for discovery of new music. Participant 12 noted:

Participant 12: *"It pushed for a lot of older music that I had never listened to before that was all in the same ball park and I really really enjoyed it. I think like, a whole bunch of Cool 'N the Gang. I had never listened to it before, but because I was listening to Stevie Wonder and it was like, "maybe you'll like this" and I loved it."*

This result was somewhat surprising because the interface was designed to refine recommendation, but it was encouraging that the interface was useful for discovery as well as refinement.

6.4.2.4 Fitting Music to Context

We were also surprised that there was no apparent relationship between correctly predicting the participants' context and their ratings of the playlists created for that context. While reviewing the interviews, I looked for possible explanation as to why this might be. One possible explanation is that we simply did not correctly capture what was important about the participants' context. In the interviews, we asked specifically about this and participants held mixed views about what else could have improved their recommendations. While some participants felt that the information gathered was sufficient, several participants noted the importance of the co-presence of friends. Two of the participants created playlists specifically for the time they spent with their significant other and another participant mentioned:

Participant 1: *"That would be really good, if you had section for different people you were listening to it with. That would be a great extra category, or way to define playlists."*

Another theme that I encountered was that different contexts suggest more or less specificity for the music required to accommodate them. For example, the participants often felt that music for studying had to be quite specific to achieve the balance of focus required. However, activities such as "hanging out" required far less specific music. Participant 1 noted this when stating:

Participant 1: "*It didn't matter as much with studying because it was such a specific playlist, that I was fine with all that stuff being on there. Within studying it's pretty much like, I'm into studying or I'm not into studying. No matter what I want study music as opposed to hanging out. There's all types of different hanging out. So, the sliders are much more useful for more general playlists that I want to tailor and get a specific feeling. As opposed to a studying playlist is just a studying playlist and it's gonna be the same music."*

Based on these comments it seems that mood and location were useful starting points for providing context-aware recommendations, but a richer description would include social factors as well. However, it also seems clear that depending on the context, the additional amount of refinement from contextual might be unnecessary.

6.5 Discussion and Future Work

Based on the findings it seems clear that Harmonizer was able to provide a unique way to create music recommendations that fit well with how the participant's viewed music. One of the key ways this was accomplished was by providing a language that both the user and the system could respond to. By providing seed songs for the participant to interact with they could infer how adjusting these songs would influence the results. Furthermore, it also provides a language that the computer can respond to and the weighting technique applied to the tags was a simple enough algorithm that it could run in real time without delaying the production of the recommendations. For the purposes of this study, it was necessary that we made use of the available tag data as a user generated language to filter against, but in principle this same technique could be applied to recommender systems that rely on feature analysis of songs such as the work by (Feng, Zhuang, & Pan, 2003) that relies on computational media aesthetics.

While the findings from this work do provide concrete insights, they also illuminate areas that require further exploration. Principle among these was the system's inability to produce recommendations that the participants viewed as more appropriate than the control interface. The interviews shed some light on this, but it is interesting that this finding is consistent with the interface evaluation, which showed that the Graded—which was used as an inspiration for the Harmonizer interface—was not the preferred interface for providing appropriate recommendations. While this interface design was successful in key ways, it seems that it could be improved to encourage the sense that it creates appropriate recommendations. To accomplish this, we could draw on Participant 6's suggestion that the interface should include a way to express not just how you feel but how you want to feel. An important consideration of this finding is that perhaps current affective computing recommenders are drawing on the wrong component to produce recommendations. Instead of trying to accurately detect emotion, this work seems to suggest that if the goal is creating appropriate recommendations, then predicting the desired emotion may be more beneficial to users than predicting their current emotional state.

Another important consideration this research suggests is the match between granularity and interface design. As we noted in the findings, there was no clear consensus on the right level of granularity for filtering. We choose songs as the unit because this was consistent with the findings of our previous study, but this is an area that could use further exploration. The consideration here is also one between ease of use and specificity. This research may be seen as suggesting that the tradeoff is largely idiosyncratic, however we do feel that a deeper exploration of this topic may illuminate principles that could be generalized across future interfaces.

6.6 CONCLUSIONS

In this chapter I have presented the findings from a field deployment with a Relating Interface called Harmonizer (research goal 6-1). Our evaluation of this system suggests that the interface was useful in improving participants' enjoyment of the recommended songs more than songs for their context alone. Use of the interface however, did not influence participants rating of the appropriateness of the music. Participants also found the interface to be useful for refining and discovering music for their context. These findings indicate that providing participants with an interface to personalize their recommendations for their context is a promising technique to improve the design of future context-aware recommender systems (research goal 6-2). However, accomplishing this relies on carefully choosing the language used by the interface such that it is easily interpreted by user. Thereby providing a language to negotiate the relationship between their context and the recommendations. The findings that participants preferred the music recommendations when using the Harmonizer interface contributes evidence toward my thesis statement that Relating Interfaces will provide recommendations that better accommodate the user context. These findings contribute empirical results supporting the utility the Relating Interfaces.

Chapter 7. CONCLUSION

Now that I have presented the body of work that conducted in my dissertation, in this chapter I will revisit the questions that motivated this work and the conclusions for each of the studies presented. I will then discuss the implications and limitations of this work.

7.1 SUMMARY OF CONTRIBUTIONS

In the introduction, I began by posing the question, "how can the user's meaning be conveyed to the technologies we design?" To answer this question, I conducted research on the practices of context-aware system designers (Chapter 3). By analyzing the artifacts and practices of context-aware designers, I provided insight into how designers view context. This analysis demonstrated that designers' views evolve as they seek to satisfy five concerns in their work: *users*, *context*, *form*, *interaction*, and *implementation*. By addressing these concerns the designers' understanding of context influences how they *frame* a design space, *encode* the contextual components of the system into a vocabulary, use this encoded vocabulary to *unify* a solution, and then *evaluate* the solution in terms of the codes. The process relies on the designer creating multiple representations of the context that are evaluated in conjunction with different concerns and by different stakeholders.

In the Chapter 4, I turned my attention to an important stakeholder in the design process—the users. To explore how users view context, I conducted a lab study in which participants were asked to discuss context-specific music recommendations with Reflektor, a chat-client music recommender. The analysis of the participants' conversations while using Reflektor revealed the importance of establishing common ground when determining context-specific music recommendations. To establish common ground, participants drew on several resources including the sonic properties of the music, its cultural significance, and the modes of interaction that the music enabled. By identifying these practices as being key to creating socially situated music recommendations, the work in Chapter 4 provided useful insights into a previously underexplored area of context-aware recommendation research.

In Chapter 5, we took the insights gained from the lab study and used them as the basis for the design of three interfaces for producing context-aware music recommendations. The strategies used by the participants in the lab study represented ways to establish the relevance of music for a given context. My collaborator and I designed three interfaces which we felt would enable, in principle, the strategies the participants engaged in during the lab study. We then created videos demonstrating how the interfaces could work and conducted an online study in which we asked participants to rate the interfaces for ease of use, specificity, appropriateness and expressiveness. They were also asked to choose their preferred interface and provide an explanation for their choice. The participants' ratings, preferred interface, and explanations, suggested that they valued interfaces that enabled personalizing recommendations while being easy to use. The participants expressed a clear preference for the interface that allowed participants to suggest graded criteria when creating recommendations because it provided a convenient tradeoff between enabling refinement, but not requiring too much adjustment to create useful recommendations. This interface evaluation contributed insights into how users perceive the efficacy of recommender systems as well as the tradeoffs between the principles discovered in Chapter 4.

In Chapter 6, I further explored how technology can help users negotiate the relationship between their context and what music is appropriate for that context by drawing on the insights from in the interface evaluation study. To evaluate these insights, my collaborator and I designed and developed a novel recommender interface called Harmonizer. The evaluation of Harmonizer suggested that providing an interface that allowed users to refine their music selections for their context was useful in improving participants' enjoyment of the recommended songs more than songs for their context alone. Use of the interface however, did not influence participants rating of the appropriateness of the music. Participants also found the interface to be useful for refining and discovering music for their context. These findings indicate that providing participants with an interface to personalize their recommendations for their context is a promising technique to improve the design of future context-aware recommender systems. However, accomplishing this relies on carefully choosing the language used by the interface such that it is easily interpreted by the user. By doing so, Relating Interfaces provide users a language to negotiate the relationship between their context and recommendations.

7.2 IMPLICATIONS

The contributions summarized above suggests several important questions that I will address in this section. This dissertation explored how interfaces can accommodate the relationship between the users' context and how the system responds. This presumes that it is possible to accommodate this relationship, but as was discussed in the lab study from Chapter 4, factors outside the control of the interface influence its ability to respond. Therefore, how should the system behave when it can't accommodate the user's views? Furthermore, when *shouldn't* technology accommodate the user's perspective? Another important consideration, is how should designers seek to accommodate perspective and what methods and tools are available to support these considerations? Finally, it is also important to consider the implications for how the user can communicate their lived experience and how this matches the systems' view of the world. Each of these three topics are discussed below.

7.2.1 *When Perspectives Can't or Shouldn't Be Accommodated*

Much of this work is predicated on the idea that the ability to accommodate the user is essentially a design consideration. However, this is not always the case. As we saw in the Reflektor study, one central challenge to producing recommendations for the participants was the limitation of available metadata. Specifically, the available metadata was not structured in a way that enabled the recommender to respond appropriately. While metadata can be structured in any number of ways, data that reflects the user's view may not be available or may be structured in a way that is contradictory to their views. This is particularly pronounced in systems that rely on user generated data sets as was the case with Reflektor and Harmonizer. These systems both hoped to draw on the "wisdom of crowds" (Kittur & Kraut, 2008) by using community applied tag data as the source when producing recommendation. While tag datasets or "folksonomies", have been found to be a beneficial source of information to draw on for the creation of recommendations (Vig et al., 2012), they can be applied inconsistently or for purposes unrelated to organizing and describing the item being tagged (Wetzker, Bauckhage, Zimmermann, & Albayrak, 2010). Because of this, spurious and unintended relationships can arise between songs which may impact the ability of the system to accommodate the user's intentions. To resolve this mismatch, better real time feedback could be incorporated into the systems design so that users are clear on how the metadata impacts the system's response. For example, with Harmonizer we chose the seed songs for the participants and did not provide feedback on how those songs influenced the recommendations. To improve the user's understanding of how the various seed songs influence the recommendations, future designs could let the user pick seed songs or give them the chance to select any number of songs. Furthermore, simply showing them the recommendations that are being produced would encourage the user's understanding of how the seed songs—as proxies for the metadata—influence the recommendations.

This mismatch between intentions and system responses is not entirely the result of the availability or structure of metadata. As context-aware systems are expected to infer more personal components of our life, such as mood, we can expect additional mismatches to occur. This mismatch is not limited to context-aware systems; it can be viewed as another representation of the social-technical gap which has been identified as a key challenge in HCI (Ackerman, 2000). In context-aware system design, this gap is exacerbated by the mismatches that results from the systems' ability to detect or infer the world and then respond accordingly. The challenge of inferring the world has been approached by the work on mediating (Dey et al., 2002) and indexical (Rantanen, 2010) interfaces. The work on the Harmonizer system provides insight into how users would like systems to capture and respond to their context, but the findings suggests that the outcome did not always match the user's expectation. This was particularly clear in how users felt that the system should respond to a desired mood instead of their current mood. This finding is consistent with prior work on affective computing systems (MacLean, Roseway, & Czerwinski, 2013) and presents an ongoing challenge for how to interpret affect and how to respond to what is interpreted. With the Harmonizer system, we intentionally left the determined mood and location open to the user's interpretation. The aim was that by doing so we could provide a more personalized system response instead of relying on our predetermined mapping from emotion to song suggestions as is found in prior work (Bauer et al., 2011; Healey et al., 1998). By providing a mechanism to negotiate a personalized system response, Relating Interfaces present an opportunity to bridge the social-technical gap created by context-aware systems.

The orientation toward negotiating personalized meaning in context-aware systems suggests an additional opportunity to provide more transparency in the mapping the system provides. This approach is more consistent with the work on intelligibility in context-aware systems (B. Y. Lim & Dey, 2011, 2013) or work visualizing uncertainty (Kay, Kola, Hullman, & Munson, 2016). The work by Kay et al. is of relevance because it aims to visualize the amount of uncertainty rather provide a description of the outcome. Context-aware systems provide an inexact detection of the world and the information used to respond provides an imperfect match, but including a mechanism that allows users to see the degree of uncertainty in how the world is detected or how the system can respond seems like a promising approach. More specifically, the system should help to visualize the uncertainty that exists between what is detected in the world and how the system could respond in a holistic way. This approach could draw on the work of worn media (Ikemiya & Rosner, 2013), which draws attention to issues of incompleteness, impermanence and imperfections. However, rather than orienting toward material differences, this work draws attention to the incompleteness that exists when designers attempt to unify the components of a solution. By calling attention to gaps that may appear between these components, Relating Interfaces can act as a mechanism to allow users to connect the disparate pieces of their context, views and the system's ability to respond according to their intentions.

In this dissertation, I assumed that it is beneficial to accommodate the user's perspective, but it is worth considering when systems should not provide the ability to account for the user's perspective. The goal of relating interfaces is to accommodate the relationship between the user's context and how the system responds. This was explored through music recommendations, but it could be extended to a variety of topics. While the selection of other forms of media based on context is likely to be unproblematic, Relating Interfaces could pose issues for how systems interpret or respond to health and wellness information. Work in the field of personal informatics has exposed the value of collaboration between health care providers and patients who track chronic diseases (C. Chung et al., 2016). Similarly, work on the lived experience of personal informatics has noted that a barrier to tracking is knowing what and how to track (Epstein et al., 2015). In both of these cases the input of professionals is valuable in supporting the health of the person tracking how contextual factors influence their health. Designers that hope to pursue interface design that accommodates the user's perspective should be careful that the system is helpful in balancing the user's views without violating the professional caregiver's views on how and how best to track their lives.

7.2.2 *Methods to Support User Perspectives*

While much of the work presented in this dissertation focuses on interface design, this work also opens opportunities for design methods. The model of context-aware design practice presented in Chapter 3 suggests that designers engage in the process of framing a design space, encoding notions of context that reside within that space, unifying a design by balancing the relevant constraints pertinent to the system and then evaluating the system in terms of the encoded notion of context. This model creates opportunities for designers interested in exploring relating interfaces for their work. One opportunity comes from exploring how context is framed. Relating interfaces should help users negotiate the relationship between their context and the behavior of the system. How context is framed and how that framing influences implementation creates an opening for design methods that focus on detecting context. With location for example, a number of methods for detecting location have been suggested all with relative merits (Varshavsky & Patel, 2009), but location need not be derived solely from detection. Dobson's work on the subtleties of location suggests alternate formulations for location that draw on the user's experience in addition to detected location (Dobson, 2005). Dobson created a taxonomy of ways that an individual's location can be determined. At the top of his taxonomic hierarchy, Dobson suggests that locations can be broken into three categories: known, approximate, and negative. For example, you might know that the user is at work (known), that they are on their way to work (approximate), or that they are not at work (negative). Furthermore, Dobson's taxonomy distinguishes between knowing someone is at work (named space) from knowing his or her exact location at work (absolute position). This is only a small subset of terms provided by Dobson to describe location, but it helps demonstrate alternate framings for location—a widely used component of context—which require no detection and can determined by the views of the user.

Drawing on the example of location illustrates a larger design consideration: how can the designer inform their understanding of how context is determined? If Relating Interfaces are expected to surface mismatches between how context, as determined by the system, is represented to the user then developing empathy for the sensors that detect context is a relevant concern to designers. By developing empathy for the sensors that detect or determine context, designers can better unify design consideration. One approach that could be of interest comes from philosopher Nagel who posed the question: 'What is it like to be a bat" (Nagel, 1974). This was a thought experiment to explore the challenge of reducing conscious experience from the lived experience. We believe that bats do have some form of consciousness, but the way their use of echolocation to perceive the world renders that form of consciousness is totally alien to us. While I am not arguing that computers are conscious, I do believe that adopting a "sensors view" of the world could provide designers with a better understanding of how the systems they design are situated in the world and therefore how they can better accommodate the users they are situated with. This method could be viewed as similar to the "Empathy Tools" IDEO method card ("IDEO: IDEO Method Cards: 51 Ways to Inspire Design.," 2003), which can be used by designers to build empathy with users. The importance difference is that this method would instead focus on building empathy with the technology and its capabilities rather than the users themselves.

7.2.3 *Supporting the Lived Experience*

Finally, it is important to consider how Relating Interfaces can support the lived experience of the users that employ them. Much of the work in this dissertation assumed that the language that users rely on to express the relationship between context and constructs can be derived from their conversations. Language is rich system used for meaning making, but it is clearly not the only factor. As was noted in Chapter 6, there are parallels between Relating Interfaces and traditional direct manipulation interfaces (Shneiderman, 1997). By drawing on the language used to discuss context, we hoped to improve on direct manipulation interfaces by providing an interface that allowed users to extend meaning beyond the computer and into the world. However, drawing on language alone may not be sufficient to capture users' views on context. Instead, it is worth considering if user views are embedded in their lived experience in a way that cannot be captured by language alone.

The Harmonizer system provided an interface that sought to conceptually blend (Fauconnier & Turner, 2008) the relationship between user's context and the related music. This was never explicitly stated, but by drawing on theories of mental models for categorization (Lakoff, 1999) we hoped to provide a way to anchor the user's understanding of their context and its relationship to the music. The outcome of the lab study provided useful insights into the design of future systems, but drawing on the behaviors of participants in natural settings likely would have revealed additional strategies that individuals engage in when establishing the relationship between their context and what music is appropriate. Of particular relevance are the material or social constraints that exist when participants determine the relationship between their context and the system's response. In the lab study the participants explored the relationship between music and context in a hypothetical space. This allowed them to elaborate a scenario where any activity, social circumstance, and ambience was equally available. In the real world, this is never the case. By analyzing the constraints that exist in real-world settings, we can see how the world influences their thoughts around context.

Work on material anchors and conceptual blending (Hutchins, 2005) has provided useful insights into cognitive work and provides a valuable lens to observe embodied practice. The research in this dissertation demonstrates the value of orienting to the cognitive work done by designers to represent and communicate context. It also demonstrates the importance of understanding users' views on context and how interfaces can be designed to accommodate these views. Moving forward, drawing on prior work on material anchors could provide a useful lens to extend the work of this dissertation to methods for analyzing field work. Doing so would improve our understanding of user views on how context is framed and encoded in the world. This dissertation therefore presents an important first step in developing theory and practice around design for context-aware systems to accommodate user's lived experience.

7.3 CONCLUSION

In this section I laid out the main contributions for each chapter in this dissertation. I then discussed the relevant implication for the design of future Relating Interfaces including the importance of emphasizing the imprecise match between detecting and responding to the user's context. I also outlined the importance of not accommodating the user's perspective in all circumstances. I then discussed the importance of adopting methods to better frame designers' views on context. I concluded with a discussion of opportunities for exploring the lived experience of users and how this could contribute to methods and theory for the design of context-aware systems. Taken together, these implications help to establish a direction and the limitations of future work on communicating meaning to context-aware systems.

BIBLIOGRAPHY

Abowd, G. D. (2012). What next, ubicomp?: celebrating an intellectual disappearing act. *Proceedings of the 2012 ACM Conference on …*.

Abowd, G. D. (2016). Beyond Weiser: From Ubiquitous to Collective Computing. *Computer*, *49*(1), 17–23. http://doi.org/10.1109/MC.2016.22

Ackerman, M. (2000). The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human-Computer Interaction*, *15*(2), 179–203. http://doi.org/10.1207/S15327051HCI1523_5

Adomavicius, G., & Tuzhilin, A. (2015). Context-aware recommender systems. *Recommender Systems Handbook, Second Edition*, 191–226. http://doi.org/10.1007/978-1-4899-7637-6_6

Agre, P. E. (1997). *Computation and Human Experience*. (Cambridge University Press, Ed.)*Meta*. New York, NY, USA.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *Submitted to Journal of Statistical Software*, *67*(1), 51. http://doi.org/10.18637/jss.v067.i01

Bauer, J. S., Jansen, A., & Cirimele, J. (2011). MoodMusic: A Method for Cooperative, Generative Music Playlist Creation. In *Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology - UIST '11 Adjunct* (p. 85). New York, New York, USA: ACM Press. http://doi.org/10.1145/2046396.2046435

Bauer, J. S., Newman, M. W., & Kientz, J. A. (2014). Thinking About Context: Design Practices for Information Architecture with Context-Aware Systems. In *Proceedings of the 2014 iConference*. Berlin, Germany.

Baumann, S., Jung, B., Bassoli, A., & Wisniowski, M. (2007). BluetunA: let your neighbour know what music you like. *Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems*, *2*, 1941–1946. http://doi.org/10.1145/1240866.1240929

Baur, D., Hering, B., Boring, S., & Butz, A. (2011). Who needs interaction anyway: exploring mobile playlist creation from manual to automatic. *Proceedings of the 16th International …*, 291–294. http://doi.org/http://doi.acm.org/10.1145/1943403.1943448

Beyer, H., & Holtzblatt, K. (1999). Contextual design. *Interactions*, *6*(1), 32–42. http://doi.org/10.1145/291224.291229

Bickmore, T. W., & Cassell, J. (2001). Relational agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01* (pp. 396–403). New York, New York, USA: ACM Press.

Bickmore, T. W., Mauer, D., & Brown, T. (2009). Context Awareness in a Handheld Exercise Agent. *Pervasive and Mobile Computing*, *5*(3), 226–235. http://doi.org/10.1016/j.pmcj.2008.05.004

Blackwell, A., Whitley, K., Good, J., & Petre, M. (2001). Cognitive factors in programming with diagrams. *Artificial Intelligence Review*, *15*(1–2), 95–114.

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2005). Affect: from information to interaction. In *CC '05 Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility* (pp. 59–68).

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *International Journal of Human-Computer Studies*, *65*(4), 275–291.

Bowen, G. a. (2008). Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research*, *8*(1), 137–152. http://doi.org/10.1177/1468794107085301

Brown, B., Sellen, A. J., & Geelhoed, E. (2001). Music Sharing as a Computer Supported Collaborative Application. In *Proceedings of Ecscw 2001* (pp. 179–198).

Buchenau, M., & Suri, J. F. (2000). Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 424–433). New York City, New York, United States: ACM.

Cacioppo, J. (2000). The psychophysiology of emotion. *Handbook of Emotions*.

Card, S. K., Newell, A., & Moran, T. P. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: L. Erlbaum Associates Inc.

Chung, C., Dew, K., Cole, A., Zia, J., Fogarty, J., Kientz, J. A., & Munson, S. A. (2016). Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient Generated-Data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (Vol. 12, pp. 770–786). http://doi.org/10.1145/2818048.2819926

Chung, E. S., Hong, J. I., Lin, J., Prabaker, M. K., Landay, J. A., & Liu, A. L. (2004). Development and evaluation of emerging design patterns for ubiquitous computing. In *DIS '04 Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 233–242). Cambridge, MA, USA: ACM.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to Discourse. *Cognitive Science*, *13*(2), 259–294. http://doi.org/10.1207/s15516709cog1302_7

Clore, G. L., & Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 628–642). Guilford Press.

Clore, G. L., & Palmer, J. E. (2009). Affective Guidance of Intelligent Agents: How Emotion Controls Cognition. *Cognitive Systems Research*, *10*(1), 21–30. http://doi.org/10.1016/j.cogsys.2008.03.002

Crawford, E. (2015). *2015 Nielsen Music U.S. Report*.

Cross, N. (1982). Designerly ways of knowing. *Design Studies*, *3*(4), 221–227. http://doi.org/10.1016/0142-694X(82)90040-0

Cross, N. (2001). Designerly Ways of Knowing: Design Discipline Versus Design Science. *Design Issues*, *17*(3), 49–55. http://doi.org/10.1162/074793601750357196

Crossen, A., Budzik, J., & Hammond, K. J. (2002). Flytrap: intelligent group music recommendation. *Proceedings of the 7th International Conference on Intelligent User Interfaces - IUI '02*, 184. http://doi.org/10.1145/502716.502748

Crozier, W. R. (1997). Music and social influence. In D. J. Hargreaves & A. C. North (Eds.), *The social psychology of music* (pp. 67–83). New York, New York, USA: Oxford University Press.

Damasio, A. R. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. Mariner Books.

Darwin, C., & Darwin, F. (1872). *The Expression of the Emotions in Man and Animals*. Cambridge University Press.

Davidoff, S., Lee, M. K., Dey, A. K., & Zimmerman, J. (2007). Rapidly exploring application design through speed dating. In *Ubicomp 2007, Proceedings of the 9th international conference on Ubiquitous computing,* (p. 429–446,). Innsbruck, Austria. http://doi.org/10.1007/11853565_2

Davidson, J. W. (2004). Music as Social behavior. In E. Clarke & N. Cook (Eds.), *Empirical Musicology: Aims, Methods, Prospects* (pp. 57–76). New York, New York, USA: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780195167498.003.0004

Denora, T. (1999). Music as a technology of the self. *Poetics*, *27*(1), 31–56. http://doi.org/10.1016/S0304-422X(99)00017-0

DeNora, T. (2000). *Music in everyday life*. Cambridge University Press.

DeNora, T. (2005). Music and Social Experience. In M. D. Jacobs & N. Weiss Hanrahan (Eds.), *The Blackwell Companion to the Sociology of Culture, Blackwell* (Jacobs M, pp. 147–159). Malden, MA: Blackwell Publishing Ltd.

Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, 4–7. http://doi.org/10.1007/s007790170019

Dey, A. K., & Abowd, G. D. (2000). Towards a better understanding of context and context-awareness. In *CHI 2000 workshop on the what, who, where, when, and how of context-awareness* (Vol. 4, pp. 1–6). Citeseer.

Dey, A. K., Abowd, G. D., & Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, *16*(2), 97–166.

Dey, A. K., Hamid, R., Beckmann, C., Li, I., & Hsu, D. (2004). a CAPpella. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04* (pp. 33–40). New York, New York, USA: ACM Press. http://doi.org/10.1145/985692.985697

Dey, A. K., & Mankoff, J. (2005). Designing mediation for context-aware applications. *ACM Transactions on Computer-Human Interaction*, *12*(1), 53–80. http://doi.org/10.1145/1057237.1057241

Dey, A. K., Mankoff, J., Abowd, G., & Carter, S. (2002). Distributed mediation of ambiguous context in aware environments. *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology UIST 02*, *4*(2), 121–130. http://doi.org/10.1145/571985.572003

Dobson, S. (2005). Leveraging the subtleties of location. In *sOc-EUSAI 2005 Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies* (pp. 189–193). New York, NY, USA.

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing*, *8*(1), 19–30. http://doi.org/10.1007/s00779-003-0253-8

Dow, S. P., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J. D. J. D., & Gandy, M. (2005). Wizard of Oz Support throughout an Iterative Design Process. *IEEE Pervasive Computing*, *4*(4), 18–26. http://doi.org/10.1109/MPRV.2005.93

Dow, S. P., Saponas, T. S. T., Li, Y., & Landay, J. A. (2006). External representations in ubiquitous computing design and the implications for design tools. In *DIS '06 Proceedings of the 6th conference on Designing Interactive systems* (pp. 241–250). University Park, PA, USA: ACM.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3), 169–200.

Epstein, D. A., Ping, A., Fogarty, J., Munson, S. A., Science, C., & Design, H. C. (2015). A Lived Informatics Model of Personal Informatics. In *UbiComp '15* (pp. 731–742). http://doi.org/10.1145/2750858.2804250

Fauconnier, G., & Turner, M. (2008). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.

Feng, Y., Zhuang, Y., & Pan, Y. (2003). Music information retrieval by detecting mood via computational media aesthetics. *Proceedings - IEEE/WIC International Conference on Web Intelligence, WI 2003*, 235–241. http://doi.org/10.1109/WI.2003.1241199

Fouse, A., Weibel, N., Hutchins, E., & Hollan, J. D. (2011). ChronoViz: a system for supporting navigation of time-coded data, 299-299-304–304. http://doi.org/10.1145/1979602.1979706

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, N.J.: Prentice-Hall.

Gaver, W. (2011). Making Spaces: How Design Workbooks Work. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, 1551–1560. http://doi.org/10.1145/1978942.1979169

Gaver, W., Sengers, P., Kerridge, T., & Kaye, J. (2007). Enhancing ubiquitous computing with user interpretation: field testing the home health horoscope. In *factors in computing* (pp. 537–546). ACM.

Gibbs, J. (1972). *Social Theory Construction*. Dryden Press.

Gonzalo-Alonso, J., De Juan, P., Garcí-A-Hortelano, E., & Iglesias, C. Á. (2009). A hybrid collaborative filtering system for contextual recommendations in social networks. In *Proceeding DS '09 Proceedings of the 12th International Conference on Discovery Science* (Vol. 5808 LNAI, pp. 393–400). http://doi.org/10.1007/978-3-642-04747-3_31

Goodman, E., & Wakkary, R. (2011). Understanding Interaction Design Practices. In *CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1061–1070). New York, NY, USA: ACM Press. http://doi.org/10.1145/1978942.1979100

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), 606–633. http://doi.org/10.1525/aa.1994.96.3.02a00100

Goodwin, C., & Goodwin, M. H. (1992). Assessments and the construction of context. In *Rethinking Context: Language as an Interactive Phenomenon* (pp. 147–189). http://doi.org/10.2307/2074658

Goodwin, C., & Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology*, *19*(May), 283–307.

Greenberg, S. (2001). Context as a dynamic construct. *Human–Computer Interaction*, *16*(2), 257–268.

Grudin, J. (2005). Three Faces of Human – Computer Interaction. *IEEE Annals Of The History Of Computing*, (October-November), 46–62. http://doi.org/Article

Grudin, J. (2012). A moving target: The evolution of human-computer interaction. In *The Human-Computer Interaction Handbook–Fundamentals, Evolving Technologies, and Emerging Application* (3rd ed., pp. 1–24). http://doi.org/10.1201/b11963-1

Han, B. J., Rho, S., Jun, S., & Hwang, E. (2010). Music emotion classification and context-based music recommendation. *Multimedia Tools and Applications*, *47*(3), 433–460. http://doi.org/10.1007/s11042-009-0332-6

Hargreaves, D. J., & North, A. C. (1999). The Functions of Music in Everyday Life: Redefining the Social in Music Psychology. *Psychology of Music*, *27*, 71–83. http://doi.org/10.1177/0305735699271007

Harrison, S., Tatar, D., & Sengers, P. (2007). The three paradigms of HCI. In *alt.CHI*.

Hartson, H. R., & Hix, D. (1989). Toward empirically derived methodologies and tools for human-computer interface development. *International Journal of Man–Machine Studies*, *31*(4), 477–494.

Healey, J., Picard, R. W., & Dabek, F. (1998). A New Affect-Perceiving Interface and Its Application to Personalized Music Selection. *In Workshop on Perceptual User Interfaces,(San Francisco)*.

Hollan, J. D., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transaction on Computer Human Interaction*, *7*(2), 174–196. http://doi.org/http://doi.acm.org/10.1145/353485.353487

Hong, J., & Landay, J. (2001). An infrastructure approach to context-aware computing. *Human-Computer Interaction*, *16*(2), 287–303.

Höök, K., & Löwgren, J. (2012). Strong concepts. *ACM Transactions on Computer-Human Interaction*, *19*(3), 1–18. http://doi.org/10.1145/2362364.2362371

Hutchins, E. (1995). Cognition in the Wild. *MIT Press*. MIT: Cambridge. http://doi.org/10.1023/A:1008642111457

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics*, *37*(10), 1555–1577. http://doi.org/10.1016/j.pragma.2004.06.008

Hutchins, E., & Hollan, J. D. (1985). Direct manipulation interfaces. *Human-Computer Interaction*, *1*(4), 311–338. http://doi.org/10.1207/s15327051hci0104_2

IDEO: IDEO Method Cards: 51 Ways to Inspire Design. (2003). Palo Alto, CA, USA: IDEO.

Ikemiya, M., & Rosner, D. K. (2013). Broken probes: toward the design of worn media. *Personal and Ubiquitous Computing*. http://doi.org/10.1007/s00779-013-0690-y

Kay, M., Kola, T., Hullman, J. R., & Munson, S. A. (2016). When (ish) is My Bus?: User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 5092–5103. http://doi.org/10.1145/2858036.2858558

Khetarpaul, S., & Gupta, S. (2012). Mining GPS traces to recommend common meeting points. *Proceedings of the 16th International Database Engineering & Applications Sysmposium*. http://doi.org/10.1145/2351476.2351497

Kirk, D. S., Durrant, A., Wood, G., Leong, T. W., Wright, P., & Tyne, N. (2016). Understanding the Sociality of Experience in Mobile Music Listening with Pocketsong. In *DIS 2016* (pp. 50–61). Brisbane, QLD, Australia. http://doi.org/10.1145/2901790.2901874

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 37–46. http://doi.org/10.1145/1460563.1460572

Kjeldskov, J., & Paay, J. (2006). Indexical interaction design for context-aware mobile computer systems. In *Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments - OZCHI '06* (p. 71). New York, New York, USA: ACM Press. http://doi.org/10.1145/1228175.1228191

Kukka, H., & Patino, R. (2009). UbiRockMachine: A Multimodal Music Voting Service for Shared Urban Spaces. In *Proceedings of the 8th International Conference on Mobile and*

Ubiquitous Multimedia - MUM 09 (pp. 1–8). Cambridge, UK. http://doi.org/10.1145/1658550.1658559

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. Language* (Vol. 64). The University of Chicago Press.

Lakoff, G. (1999). Cognitive Models and Prototype Theory. In E. Margolis & S. Laurence (Eds.), *Concepts: Core Readings* (2nd ed., p. 652). Cambridge, MA, USA: MIT Press.

Landay, J. A., & Borriello, G. (2003). Design patterns for ubiquitous computing. *Computer*, *36*(8), 93–95. http://doi.org/10.1109/MC.2003.1220589

Leong, T. W., & Wright, P. C. (2013). Revisiting social practices surrounding music. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (p. 951). http://doi.org/10.1145/2470654.2466122

Leung, L. (2015). Using Tablet in Solitude for Stress Reduction. *Computers in Human Behavior*, *48*(C), 382–391. article. http://doi.org/10.1016/j.chb.2015.01.068

Lewis, C., & Rieman, J. (1993). *Task-centered user interface design*. *A Practical Introductio*.

Li, Y., Hong, J. I. J., & Landay, J. A. (2007). Design challenges and principles for Wizard of Oz testing of location-enhanced applications. *IEEE Pervasive Computing*, *6*(2), 70–75. http://doi.org/10.1109/MPRV.2007.28

Li, Y., Hong, J. I., & Landay, J. A. (2004). Topiary: a tool for prototyping location-enhanced applications (pp. 217–226). Santa Fe, NM, USA: ACM.

Li, Y., & Landay, J. A. (2008). Activity-based prototyping of ubicomp applications for long-lived, everyday human activities (pp. 1303–1312). Florence, Italy: ACM.

Likamwa, R., Yunxin, L., Lane, N., & Zhong, L. (2013). MoodScope : Building a Mood Sensor from Smartphone Usage Patterns. In *MobiSys' 2013* (Vol. 39, p. 2011).

Lim, B. Y., & Dey, A. K. (2011). Investigating Intelligibility for Uncertain Context-aware Applications. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 415–424. http://doi.org/10.1145/2030112.2030168

Lim, B. Y., & Dey, A. K. (2013). Evaluating Intelligibility Usage and Usefulness in a Context-Aware Application, 92–101. http://doi.org/10.1007/978-3-642-39342-6_11

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, 2119–2129. http://doi.org/10.1145/1518701.1519023

Lim, Y.-K., Stolterman, E., & Tenenberg, J. J. (2008). The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *15*(2), 1–27.

Lincoln, Y. S., Egon G. Guba, & Guba, E. G. (1985). *Naturalistic inquiry*. *Sage* (Vol. 75). BOOK, Beverly Hills, Calif.: Sage Publications, Inc.

Liu, K., & Reimer, R. A. (2008). Social playlist: enabling touch points and enriching ongoing relationships through collaborative mobile music listening. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 403–406). http://doi.org/10.1145/1409240.1409299

MacIntyre, B., Gandy, M., Dow, S. P., & Bolter, J. D. (2004). DART: a toolkit for rapid design exploration of augmented reality experiences (pp. 197–206). Santa Fe, NM, USA: ACM.

MacLean, D., Roseway, A., & Czerwinski, M. (2013). MoodWings - A Wearable Biofeedback Device for Real- Time Stress Intervention. *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '13*, 1–8. http://doi.org/10.1145/2504335.2504406

Maguire, M. (2001). Methods to support human-centered design. *International Journal of Human-Computer Studies*, *55*(4), 587–634. http://doi.org/10.1006/ijhc.2001.0503

McCarthy, J. F. E., & Anagnost, T. D. (1998). MUSICFX: an arbiter of group preferences for computer supported collaborative workouts. In *CSCW '1998 Proceedings of the 1998 ACM conference on Computer supported cooperative work* (p. 348). Philadelphia, Pennsylvania, United States: ACM.

Moggridge, B. (2006). *Designing Interactions*. *Book*. Cambridge: The MIT Press.

Morville, P., & Rosenfeld, L. (2008). *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites*. Sebastopol, CA: O'Reilly Media, Inc.

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435–450. http://doi.org/10.2307/2176743

Newman, M. W., Ackerman, M. S., Kim, J., Prakash, A., Hong, Z., Mandel, J., & Dong, T. (2010). Bringing the field into the lab: supporting capture and replay of contextual data for the design of context-aware applications. In *UIST '10 Proceedings of the 23nd annual ACM symposium on User interface software and technology* (pp. 105–108). New York, New York, USA: ACM Press. http://doi.org/10.1145/1866029.1866048

Newman, M. W., & Landay, J. A. (2000). Sitemaps, storyboards, and specifications: a sketch of Web site design practice. *... of the 3rd Conference on Designing ...*, 263–274.

Norman, D. A. (1991). Cognitive artifacts. In J. M. Carroll (Ed.), *Designing interaction: Psychology at the human-computer interface.* (pp. 17–38). New York.

O'Hara, K., Lipson, M., Jansen, M., Unger, A., Jeffries, H., & Macer, P. (2004). Jukola: Democratic Music Choice in a Public Space. In *Proceedings of the 2004 conference on Designing interactive systems processes, practices, methods, and techniques - DIS '04* (p. 145). Cambridge, MA, USA: ACM. http://doi.org/10.1145/1013115.1013136

Picard, R. W. (2000). Affective computing. *MIT Press*.

Raacke, J., & Bonds-Raacke, J. (2008). MySpace and Facebook: Applying the uses and gratifications theory to exploring friend-networking sites. *Cyberpsychology & Behavior*, *11*(2), 169–174. article.

R Core Team. (2012). *R: A language and environment for statistical computing.* Vienna, Austria.

Ragno, R., Burges, C. J. C., & Herley, C. (2005). Inferring similarity between music objects with application to playlist generation. *Proc. 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 73–80. http://doi.org/10.1145/1101826.1101840

Rantanen, M. J. (2010). Indexicality of language and the art of creating treasures. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (p. 301). New York, New York, USA: ACM Press. http://doi.org/10.1145/1753326.1753371

Rogers, Y. (2006). Moving on from Weiser's Vision of Calm Computing: engaging UbiComp experiences. *Ubicomp '06*, 404–421. http://doi.org/10.1007/11853565_24

Rosch, E. (1999). Principles of Categorization. In E. Margolis & S. Laurence (Eds.), *Concepts: Core Readings* (2nd ed., pp. 189–206). Cambridge, MA, USA: MIT Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493–502. http://doi.org/10.1037//0022-3514.57.3.493

Saffer, D. (2006). *Designing for Interaction: Creating Smart Applications and Clever Devices* (1st ed.). Berkeley CA: Peachpit Press.

Salvador, T., & Anderson, K. (2003). Practical considerations of context for context based systems: An example from an ethnographic case study of a man diagnosed with early onset alzheimer's disease. *UbiComp 2003: Ubiquitous Computing*, 243–255. http://doi.org/doi:10.1007/978-3-540-39653-6_19

Scherer, K. R. (2000). Psychological models of emotion. In *The neuropsychology of emotion.* (pp. 137–162). CHAP, New York, NY, US: Oxford University Press.

Schilit, B. N., Adams, N., & Want, R. (1994). Context-aware computing applications. In *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on* (pp. 85–90). IEEE.

Schilit, B. N., Hilbert, D. M., & Trevor, J. (2002). Context-aware communication. *Wireless Communications, IEEE*, *9*(5), 46–54.

Schmidt, A. (2000). Implicit human computer interaction through context. *Personal and Ubiquitous Computing*, *4*(2–3), 191–199.

Schön, D. A. (1992). Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems*, *5*(1), 3–14. JOUR.

Schutz, A. (1962). *Collected papers I: The Problem of Social Reality*. (M. Natason, Ed.). The Hague, The Netherlands: Martinus Nijhoff.

Sease, R., & McDonald, D. W. (2009). Musical fingerprints: Collaboration around home media collections. *Proceedings of the ACM 2009 International Conference on Supporting Group Work*, 331–340. http://doi.org/10.1145/1531674.1531724

Seeburger, J., Foth, M., & Tjondronegoro, D. (2012). The sound of music: sharing song selections between collocated strangers in public urban places. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM '12* (p. 1). New York, New York, USA: ACM Press. http://doi.org/10.1145/2406367.2406409

Sellen, A. J., Rogers, Y., Harper, R., & Rodden, T. (2009). Reflecting human values in the digital age. *Communications of the ACM*, *52*(3), 58. http://doi.org/10.1145/1467247.1467265

Sen, S., Harper, F. M., LaPitz, A., & Riedl, J. (2007). The quest for quality tags. *Proceedings of the 2007 International ACM Conference on Conference on Supporting Group Work - GROUP '07*, 361. http://doi.org/10.1145/1316624.1316678

Setten, M. Van, Instituut, T., Mcnee, S. M., & Konstan, J. A. (2005). Beyond personalization: the next stage of recommender systems research. *Human Factors*, 7500–7500. http://doi.org/10.1145/1040830.1040839

Shepherd, J., & Wicke, P. (1997). *Music and cultural theory*. Malden, MA: Polity Press.

Shneiderman, B. (1983). Direct Manipulation: A Step Beyond Programming Languages. *IEEE Computer*, *16*(8), 57–69. http://doi.org/10.1109/MC.1983.1654471

Shneiderman, B. (1997). Direct Manipulation for Comprehensible, Predictable and Controllable User Interfaces. *Proceedings of the 2Nd International Conference on Intelligent User Interfaces*, 33–39. http://doi.org/10.1145/238218.238281

Simon, R., & Fröhlich, P. (2007). A mobile application framework for the geospatial web. *16th International World Wide Web Conference, WWW2007*, 381–390. http://doi.org/10.1145/1242572.1242624

Sinha, R., & Swearingen, K. (2002). The role of transparency in recommender systems. In *CHI 02 extended abstracts on Human factors in computing systems CHI 02* (p. 830). http://doi.org/10.1145/506443.506619

Sohn, T., & Dey, A. K. (2003). iCAP. In *CHI '03 extended abstracts on Human factors in computing systems - CHI '03* (p. 974). New York, New York, USA: ACM Press. http://doi.org/10.1145/765891.766102

Sørensen, H., & Lagerl, S. (2012). The Interaction Space of a Multi-Device , Multi-User Music Experience. In *NordiCHI '12 Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design Pages 504-513* (pp. 504–513). New York, NY, USA. http://doi.org/10.1145/2399016.2399094

Sprague, D., Wu, F., & Tory, M. (2008). Music selection using the PartyVote democratic jukebox. In *Proceedings of the working conference on Advanced visual interfaces - AVI '08* (p. 433). http://doi.org/10.1145/1385569.1385652

Stolterman, E. (2008). The Nature of Design Practice and Implications for Interaction Design Research. *International Journal of Design*, *2*(1), 55–65.

Suchman, L. A. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press.

Swearingen, K., & Sinha, R. (2002). Interaction design for recommender systems. *Designing Interactive Systems*, 1–10. http://doi.org/10.1.1.15.7347

Thayer, R. E. (1978). Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion*, *2*(1), 1–34. http://doi.org/10.1007/BF00992729

Tomkins, S. S. (1984). Affect theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (Vol. 163, pp. 163–195). Erlbaum.

Truong, K., Huang, E., & Abowd, G. (2004). CAMP: A magnetic poetry interface for end-user programming of capture applications for the home. *UbiComp 2004: Ubiquitous …*.

Varshavsky, A., & Patel, S. (2009). Location in Ubiquitous Computing. In J. Krumm (Ed.), *Ubiquitous Computing Fundamentals* (pp. 285–320). CRC PRess.

Vig, J., Sen, S., & Riedl, J. (2012). The Tag Genome: Encoding Community Knowledge to Support Novel Interaction. *ACM Transactions on Interactive Intelligent Systems*, *2*(3), 1–44. http://doi.org/10.1145/2362394.2362395

Voida, A., Grinter, R. E., Ducheneaut, N., Edwards, W. K., & Newman, M. W. (2005). Listening in: practices surrounding iTunes music sharing. In *Proceedings of the 2005 Annual Conference on Human Factors in Computing Systems - CHI '05*. http://doi.org/10.1145/1054972.1054999

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, *265*(3), 94–104. http://doi.org/http://doi.acm.org/10.1145/329124.329126

Weiser, M. (1993). Some computer science issues in ubiquitous computing. *Communications of the ACM*, *36*(7), 75–84.

Weiser, M. (1994). Creating the invisible interface: (invited talk). *UIST 1994*, 1. http://doi.org/http://doi.acm.org/10.1145/192426.192428

Wetzker, R., Bauckhage, C., Zimmermann, C., & Albayrak, S. (2010). I Tag , You Tag: Translating Tags for Advanced User Models. *Third ACM International Conference on Web Search and Data Mining*, *February*, 71–80. http://doi.org/10.1145/1718487.1718497

Wyer, R. S., Clore, G. L., & Isbell, L. M. (1999). Affect and information processing. *Advances in Experimental Social Psychology*, *31*(x), 1–77.

Zheng, V. W., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative location and activity recommendations with GPS history data. *Proceedings of the 19th International Conference on World Wide Web WWW 10*, 1029. http://doi.org/10.1145/1772690.1772795