Title
Augmenting collocated interaction : the design of assistive technology to support face-to-face communication

Author
Piper, Ann Marie

Publication Date
2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Augmenting Collocated Interaction:
The design of assistive technology to support face-to-face communication

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Cognitive Science

by

Anne Marie Piper

Committee in charge:

Professor James D. Hollan, Chair
Professor Charles Goodwin
Professor William Griswold
Professor Edwin Hutchins
Professor Terry Jernigan
Professor Carol Padden

2011

The dissertation of Anne Marie Piper is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____

____
Chair

University of California, San Diego

2011

DEDICATION

For my mother, Barbara Piper.

EPIGRAPH

Deafness has left me acutely aware of
both the duplicity that language is capable of
and the many expressions the body cannot hide.
—Terry Galloway, Deaf writer and performer

I remember the first day you said my name.
I remember the first day you said you love me,
and I remember that it took a long time.
—Louise, wife of man with aphasia

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Many people contributed to the work within this thesis and my success as a graduate student. I am sincerely grateful to have worked with and learned from Jim Hollan, my advisor. From the start, he gave me the freedom to pursue my own research ideas and explore the problems I found most interesting. Jim always saw the bigger vision of our research and helped me do the same. Ed Hutchins encouraged me to think deeply about theories of distributed cognition and was patient with me as I grew as a researcher. Carol Padden has been an exceptional mentor. She shaped my perspectives on how we might design culturally appropriate communication technologies. I appreciate Bill Griswold's ability to see the technical contributions of my research and how he helped me draw out those aspects. Thank you to Chuck Goodwin for taking an interest in my work and being a constant source of positive encouragement. His excitement and thoughtfulness in this area of work is contagious. I thank Terry Jernigan for working with me to understand the complexities of aging over the lifespan and for helping me think about how we can integrate this understanding into the design of technologies for older adults.

I have benefited greatly from the support of members of the Distributed Cognition and Human-Computer Interaction Lab at UCSD. Thank you to Amaya Becvar, Gaston Cangiano, Brynn Evans, Adam Fouse, Whitney Friedman, Chris Johnson, Nancy Owens Renner, and Nadir Weibel. One of my favorite parts of graduate school was working with outstanding undergraduate students. I am thankful to have worked with Ross Campbell, Jordan Davison, Amanda Lazar, Amanda Legge, Anna Ostberg, and Matt Wilczynski, all of whom are now off doing great work of their own. Thank you, Lisa Cowan, for your encouragement throughout my advancement and defense process. It was wonderful to go through this process together. I want to thank the members of the Women in HCI @ UCSD group, who were always excited to talk about research ideas and understood the importance of forming a support network in academia.

There are several people within the broader San Diego community who played a critical role in my dissertation research. I thank Georgia Sadler for her idea behind the Shared Speech Interface project and for helping me connect with the local Deaf community. Thank you, Tress Goodwin, for offering great insights into the medical community and for making several of my research projects possible. I also want to thank Jason Berganica for offering his time and knowledge in support of my work.

I have had many outstanding mentors and collaborators over the past decade. They include Eileen O'Brien, Meredith Morris, and Terry Winograd at Stanford University, Amy Bruckman and Gillian Hayes at Georgia Tech, Kristin Alexander at Microsoft, and Jim Gray and Julie Hwang at LeapFrog. Each of these people helped shape me as a researcher, and I am grateful for their mentorship.

Finally, I thank my immediate and extended family for their continued support and encouragement in completing my degree. Above all, I am grateful for the unwavering support of my husband. I owe him my success. Thank you, Mike, for always believing in me, listening to me talk incessantly about research projects, and encouraging me to pursue the work that I love.

Chapter 3, in part, is a reprint of the material as it appears in the Proceedings of the 2006 ACM Conference on Computer-Supported Cooperative Work (CSCW). Piper, A.M., O'Brien, E., Morris, M.R., and Winograd, T. SIDES: A Cooperative Tabletop Computer Game for Social Skills Development. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is a reprint of the material as it appears in the Proceedings of the 2008 ACM Computer-Supported Cooperative Work (CSCW). Piper, A.M. and Hollan, J. Supporting Medical Conversations between Deaf and Hearing Individuals with Tabletop Displays. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in part, is a reprint of the material as it appears in the Proceedings of the 2009 European Conference on Computer-Supported Cooperative Work (ECSCW). Piper, A.M. and Hollan, J. Analyzing Multimodal Communication around a Shared Tabletop Display. The dissertation author was the primary investigator and author of this paper.

Chapter 6, in part, is a reprint of the material as it appears in the Proceedings of the 2010 ACM Conference on Computers and Accessibility (ASSETS). Piper, A.M., Weibel, N., and Hollan, J. Introducing Multimodal Paper-Digital Interfaces for Speech-Language Therapy. The dissertation author was the primary investigator and author of this paper.

Chapter 7, in part, has been submitted for publication of the material as it may appear in the ACM Transactions on Accessible Computing (TACCESS). Piper, A.M., Weibel, N., and Hollan, J. Write-N-Speak: Introducing Multimodal Paper-Digital Interfaces and an Authoring Toolkit for Speech-Language Therapy. The dissertation author was the primary investigator and author of this paper.

VITA

2004	B. S. in Computer Science, Georgia Institute of Technology
2005	M. A. in Education, Stanford University
2011	Ph. D. in Cognitive Science, University of California, San Diego

PUBLICATIONS

Piper, A.M., Weibel, N., and Hollan, J. Write-N-Speak: Introducing Multimodal Paper-Digital Interfaces and an Authoring Toolkit for Speech-Language Therapy. ACM Transactions on Accessible Computing (TACCESS), *under review*.

Piper, A.M. and Hollan, J. Supporting Medical Communication for Older Patients with a Shared Touch-Screen Computer. International Journal of Medical Informatics, *accepted*.

Piper, A.M., Weibel, N., and Hollan, J. Introducing Multimodal Paper-Digital Interfaces for Speech-Language Therapy. Proceedings of ACM Conference on Computers and Accessibility (ASSETS) 2010, 203-210.

Weibel, N., Piper, A.M., and Hollan, J. HIPerPaper: Introducing Pen and Paper Interfaces for Ultra-Scale Wall Displays. Symposium on User Interface Software and Technology (UIST) 2010 demonstration, 407-408.

Weibel, N., Piper, A.M., and Hollan, J. Exploring Pen and Paper Interaction with High-Resolution Wall Displays. Symposium on User Interface Software and Technology (UIST) 2010 poster, 455-456.

Piper, A.M., Campbell, R., and Hollan, J. Exploring the Accessibility and Appeal of Surface Computing for Older Adult Health Care Support. Proceedings of Conference on Human Factors in Computing Systems (CHI) 2010, 907-916.

Piper, A.M. Supporting Medical Communication with a Multimodal Surface Computer. Doctoral Consortium Extended Abstracts of Conference on Human Factors in Computing Systems (CHI) 2010, 2899-2902.

Piper, A.M. and Hollan, J. Analyzing Multimodal Communication around a Shared Tabletop Display. Proceedings of European Conference on Computer-Supported Cooperative Work (ECSCW) 2009, 283-302.

Piper, A.M. and Hollan, J. Tabletop Displays for Small Group Study: Affordances of Paper and Digital Materials. Proceedings of Conference on Human Factors in Computing Systems (CHI) 2009, 1227-1236.

Piper, A.M. and Hollan, J. Supporting Medical Conversations between Deaf and Hearing Individuals with Tabletop Displays. Proceedings of Computer-Supported Cooperative Work (CSCW) 2008, 147-156.

Morris, M.R., Piper, A.M., Cassanego, A., Huang, A., Paepcke, A., and Winograd, T. Mediating Group Dynamics through Tabletop Interface Design. IEEE Computer Graphics and Applications, Sept/Oct 2006, 65-73.

Piper, A.M., O'Brien, E., Morris, M.R., and Winograd, T. SIDES: A Cooperative Tabletop Computer Game for Social Skills Development. Proceedings of Computer-Supported Cooperative Work (CSCW) 2006, 1-10.

Hayes, G.R., Piper, A.M., Amar, B., Bevis, K.J., Newstetter, W., and Bruckman, A.S. Audience in Computer Learning: A Constructionist Interpretation. In the Extended Abstracts of the International Conference of the Learning Sciences (ICLS) 2004.

ABSTRACT OF THE DISSERTATION

Augmenting Collocated Interaction:
The design of assistive technology to support face-to-face communication

by

Anne Marie Piper

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2011

Professor James D. Hollan, Chair

Effective face-to-face communication with other people presents challenges for many reasons. We may be distracted, fatigued, or emotional and unable to communicate well or remember the details of a conversation. Certain contexts, such as medical and therapeutic interaction, present additional challenges to effective communication due to illness, stress, or disability. This thesis presents research on novel systems to support collocated interaction for populations who have communication challenges related to a hearing, speech, or developmental disability. I describe three research projects that involve understanding face-to-face communication needs, designing novel systems to augment interaction among collocated participants, and evaluating how these systems shape the nature of human-human interaction. First, the SIDES project introduces a cooperative tabletop computer game designed as a social skills therapy tool for children with Asperger's Syndrome, an Autism Spectrum Disorder. Children with AS often have difficulty understanding accepted social conventions, reading facial expressions, interpreting body language, and understanding social protocols. Findings indicate that cooperative tabletop computer games are a motivating and supportive tool for facilitating face-to-face interaction involving this population. Second, the Shared Speech Interface (SSI) project involves the design and evaluation of an application for an interactive multitouch tabletop display that facilitates medical communication between a Deaf patient and a hearing, non-signing physician. SSI provides Deaf individuals with a more private and independent alternative for medical communication. SSI also reshapes communication between the doctor and Deaf patient in important ways. Third, the Write-N-Speak project examines face-to-face communication for individuals with aphasia. Through a year-long field study, I understand the process of speech-language therapy for older adults with aphasia and introduce digital pen technology into this work environment. This project also involves the design and field deployment of Write-N-Speak, a programmable toolkit that allows non-technical end-users to independently create custom interactive paper materials. Through this thesis, I provide a deeper understanding of face-to-face human interaction involving critical user populations, such as children with autism, Deaf people, and older adults with aphasia. I introduce novel prototype systems that support face-to-face interaction among participants with varying abilities and examine how these systems augment collocated interaction.

Chapter 1

Introduction

Effective face-to-face communication with other people can be challenging for a variety of reasons. We may be distracted, fatigued, or emotional and unable to communicate well or remember the details of a conversation. Certain contexts present additional challenges to communication. For example, medical communication between a doctor and patient is challenging because factors such as illness, differences in knowledge levels, and limited time complicate the interaction, preventing the effective exchange of information. Now, consider the rapidly growing population of people over age 65 who arguably have the greatest health care needs, but whose age-related changes in hearing, visual, cognitive, and physical ability make face-to-face communication even more demanding (World Health Organization, 2011a). This thesis is concerned with understanding populations for whom face-to-face interaction is challenging and then designing technology-based solutions that support this interaction in context.

While face-to-face communication can be challenging for everyone, individuals with a communication related disability experience additional challenges during interactions. *Assistive technology* encompasses the design and use of computer technology to support the needs of people with disabilities. The World Health Organization (2011b) defines *disability* as "an umbrella term, covering impairments, activity limitations, and participation restrictions... Thus disability is a complex phenomenon, reflecting an interaction between features of a person's body and features of the society in which he or she lives." The word disability has a negative connotation for some communities; yet, at the same time, this classification provides access rights and government assistance to key populations. In writing this thesis I take a person-first approach (e.g., a woman with aphasia) to emphasize the individual rather than the disability. More importantly, I take an augmentative approach, which has a subtle but important difference from a deficit approach. Instead of using technology to fill a gap or compensate for one's loss, I seek to apply technology in ways that augment and enhance one's capabilities.

Research within this thesis focuses on understanding and designing assistive technology to support collocated communication involving individuals with a hearing, speech, or developmental disability. The vast majority of assistive communication devices are designed for a single user. Furthermore, these devices may not take the larger context of ongoing activity into consideration, nor accommodate the constraints and needs of the interlocutor. In this work, I examine and design for collocated communication around table and desk-like workspaces. Tables and desks are often an organizing feature of human interaction, positioning people in a face-to-face orientation while also providing a communal space for sharing artifacts. New technologies such as *surface computers* and *digital pens* transform traditional horizontal surfaces into rich interactive workspaces for collaboration and communication. These technologies can not only support collocated communication, but also present an opportunity to enable more natural forms of human-computer interaction through the use of gesture and speech input.

A core tenet of this thesis is that face-to-face human interaction is inherently multimodal: we fluidly use speech, gesture, eye gaze, and our body to interact with other people (Garfinkel, 1967; Schegloff and Sacks, 1973; Schegloff, 1984; Goffman, 1981; Kendon, 1967, 1980; Goodwin, 1981, 1986). This framing is essential when studying individuals with communication challenges and contexts where communication depends on multimodal representations of language. For example, work by Goodwin (1995; 2003a; 2004; 2006) illustrates how a man with aphasia acts as a competent speaker by orchestrating the world around him. Our activity is situated and shaped by the actions of others, the artifacts and context surrounding interaction, and cultural practices (Suchman, 1987; Hutchins, 1995). During interaction, we exploit a variety of cognitive, social, and environmental resources to achieve coordinated activity. Theories of *distributed cognition* (Hutchins, 1995; Norman, 1993; Salomon, 1993) are the foundation for how I understand human interaction in computer-mediated workspaces. Distributed cognition expands the unit of analysis beyond the individual and considers the larger system of activity, including multiple people, their actions, artifacts in the environment, cultural practices, and the transformative nature of these representations and processes over time. Additionally, the human body and material workspace are central to analysis. Interaction is fundamentally embodied, and cognition is an emergent property of interaction. Distributed cognition is an ideal theoretical framework for studying human activity in complex computer workspaces (Hollan et al., 2000; Hollan and Hutchins, 2009; Halverson, 2002).

1.1 Research Questions

At a high-level, this dissertation addresses the question ***How should we design computer systems to augment collocated interaction involving individuals with varying communication abilities?*** This question spans issues of understanding users, designing usable and useful interfaces, implementing prototypes, and evaluating the impact of these prototypes on human-human interaction. Several subquestions are important within the context of this design space.

- *If human-human interaction is inherently multimodal, how then should we design systems to augment interaction?* Multimodal systems that integrate speech- and gesture-based input are one promising technique. I hypothesize that providing users with multiple modalities for interaction with computer systems is critical in the design of workspaces that support users with varying abilities.

- *How does the introduction of novel multimodal technologies (e.g., surface computers) support interaction between participants in a collocated workspace?* One goal of system design is to support face-to-face interaction in such a way that the technology augments the person-to-person experience rather than requiring people to focus attention on the system. How might a horizontal multimodal workspace support this goal?

- *When designing these multimodal systems, which mechanisms are effective for capturing, representing, and repurposing dialogue in face-to-face communica-*

tion? In this thesis I examine techniques to reify speech on a shared multitouch display as well as capture and repurpose speech using a multimodal digital pen. One goal is to do so accurately and naturally as part of the interaction. One hypothesis is that reifying dialogue reshapes the way people interact in a conversational setting because, for example, such interfaces allow participants to review prior conversation and use it in future communication.

- *How do we evaluate these novel multimodal communication systems in ways that provide deep understandings of their impact on human interaction?* Innovation in communication technologies requires innovation in the methods we use to evaluate these systems. Theories of distributed cognition provide an ideal framework for understanding the nature of interaction through new technology systems. Specifically, distributed cognition places emphasis on the human body in interaction, making it critical to examine patterns of speech, gesture, eye gaze, and body movement. Also within this question, I hypothesize that aspects of communication, such as increased comfort, privacy, and motivation are equally important, if not more important than efficiency in interaction. With distributed cognition, understanding the culture of a user group and the tradeoffs within an actual context of use is key.

1.2 Research Approach and Contributions

Research in this thesis applies a human-centered approach that has an early emphasis on users and involves iterative design. Distributed cognition is central to how I conduct research, design systems, and evaluate their impact on interaction. In the first phase of research, I gain an understanding of user needs through theoretical and empirical foundations. This step in the design process involves surveying literature on human-human interaction and reviewing existing communication alternatives (digital and non-digital) for the population of interest. Importantly, I also gather empirical data through field observations and/or interviews with potential system users and domain experts. This phase requires establishing relationships with various user groups in the local community. For example, in this research, I partnered with local Deaf organizations, Hard of Hearing support groups, clinical training programs, and retirement communities. An initial understanding of the broader culture in which a potential solution will be positioned is established during this first phase, although this understanding evolves over the project lifecycle. Data from this phase of research contributes: (1) a better understanding of the challenges various populations experience in face-to-face communication, and (2) empirical and theoretical requirements for the design of an augmentative communication system.

The second phase of research involves the user interface design and implementation of a system that augments communication for the population of study. Augmentative system design takes the approach that the technology complements people's existing abilities rather than focusing on a disability as a deficit or gap that the system will fill. Natural user interfaces often involve input through touch or speech, and the prototype systems within this thesis exploit those input techniques. While gesture-based input has been around for several decades (e.g., Krueger et al., 1985), touch input is now pervasive because of systems like Microsoft Surface[1], Apple iPad[2], and iPhone[3]. Another growing gesture-based input technique is pen-based computing. Commercial versions of digital pen technology by Anoto[4], Livescribe[5], and LeapFrog[6] allow interaction by writing or performing pen-based gestures on paper printed with a special dot pattern. Similarly, high-quality speech recognition systems are now commercially available (e.g., Dragon NaturallySpeaking[7]). Multimodal systems that integrate gesture and speech promise greater flexibility of user input, improved system accessibility, and reduced error rates (Oviatt, 1999; Oviatt and VanGent, 1996; Oviatt, 2002). Several systems in this thesis also include multimedia output, providing redundancy in communication by simultaneously presenting visual and auditory representations of language. The main contribution of this phase of research is the introduction of novel prototype systems designed to support users with varying communication abilities.

The third phase of research involves evaluating how these prototype systems im-

[1]http://www.microsoft.com/surface
[2]http://www.apple.com/ipad
[3]http://www.apple.com/iphone
[4]http://www.anoto.com
[5]http://www.livescribe.com
[6]http://www.leapfrog.com
[7]http://www.nuance.com/dragon

pact interaction involving individuals with disabilities and their interlocutor. For each system, the goal is to support effective face-to-face communication. Specific outcomes for each system vary, however, and may include improved privacy, independence, motivation, ability to express oneself, and conversational understanding. By leveraging theories of distributed cognition, I examine the system of interaction as it relates to these outcomes. This often involves micro-analysis of interaction, which is an effective way to describe the intricacies of complex human-computer systems in action. Work within this thesis contributes to the field of human-computer interaction by demonstrating how distributed cognition facilitates deep understandings of the ways in which collocated communication systems shape interaction. System evaluations also include behavioral performance data (e.g., time to complete a task) and self-report data collected via questionnaires. Interviews with participants after they use a new system are another common source of data. Finally, this approach to research is not usually a linear process. System development often requires iterative prototyping and evaluation with users over time.

1.3 Thesis Overview

This thesis is written with two audiences in mind. The primary audience for this work is the community of human-computer interaction (HCI), computer-supported cooperative work, and assistive technology researchers who are interested in supporting interaction between collocated participants with varying abilities. Additionally, I hope this thesis will be of benefit to educators, clinicians, practitioners, and caregivers who are often responsible for the ultimate adoption and use of assistive technologies.

In Chapter 2, I describe related work that is the foundation for this thesis. This chapter reviews research on the development of interactive tabletop and desktop workspaces, including various interaction techniques. I also discuss relevant research on evaluating activity in multiuser human-computer workspaces.

Chapter 3 presents SIDES, a cooperative tabletop computer game designed as a social skills therapy tool for children with Asperger's Syndrome (AS), an Autism Spectrum Disorder. Children with AS often have difficulty understanding accepted social conventions, reading facial expressions, interpreting body language, and understanding social protocols. This chapter describes a design case study conducted over six months with a middle school social skills therapy class. Findings indicate that cooperative tabletop computer games are a motivating and supportive tool for teaching face-to-face communication skills to children with AS.

Chapter 4 describes the design and early evaluation of Shared Speech Interface (SSI), an application for an interactive multitouch tabletop display designed to facilitate medical conversations between a Deaf patient and a hearing, non-signing physician. This chapter presents a comparative laboratory evaluation that contrasts SSI with conversation facilitated by a human sign language interpreter. One main advantage of this system is that it provides Deaf individuals with a more private and independent alternative for medical communication.

Chapter 5 presents a detailed analysis of interaction through SSI. This analysis indicates that SSI allows the Deaf patient to watch the doctor when she is speaking, encourages the doctor to exploit multimodal communication, such as co-occurring gesture-speech, and provides shared access to persistent, collaboratively produced representations of conversation.

Chapter 6 addresses the design space around supporting face-to-face communication for individuals with aphasia. Aphasia, a communication condition often resulting from a stroke, is characterized by a reduced ability to understand and/or generate speech and language. A year-long field study examines the process of speech-language therapy for older adults with aphasia and introduces digital pen technology into this work environment.

Chapter 7 describes the design of Write-N-Speak, a programmable toolkit that allows non-technical end-users to independently create custom interactive paper materials for therapy. This chapter also reports on a 12-week clinical field study that evaluates the Write-N-Speak system.

Chapter 8 summarizes the work within this thesis. This chapter also describes the contributions of this thesis to the field of HCI.

Chapter 2

Related Work

This chapter describes the theoretical and empirical foundations upon which this dissertation research is based. First, I review the evolution of novel technologies to support face-to-face interaction around table and desktop surfaces. Then, I propose distributed cognition as a productive theoretical framework for research on collocated communication in interactive workspaces. I briefly review research that examines interaction around traditional tabletop workspaces. Finally, I discuss research that examines the practices and nature of interaction in sociotechnical systems involving computer display technology.

2.1 Introduction

This thesis focuses on designing computer systems and interaction techniques that support multiple people in a collocated workspace. Most often this involves technologies that support interaction around table and desk surfaces. Often included in this category are *multitouch tabletop displays*, also called "surface computers" due to the introduction of Microsoft Surface. These systems support interaction with digital materials that are projected or displayed onto a horizontal surface. Single display groupware (SDG) (Stewart et al., 1999) is the broader category of technology that supports multiple people interacting with a single computer interface in a collocated setting. It has been suggested that SDG leads to greater task engagement and activity participation by people involved in collocated computer-mediated activity (Stewart et al., 1999). I now review seminal research on horizontal interactive workspaces and various interaction techniques that influence how we design and think about cooperative computing.

2.2 Early Systems and Interaction Techniques

2.2.1 First Horizontal Computer Workspaces

Myron Krueger's VideoDesk (Krueger, 1991, 1993) was one of the first systems to explore augmented interaction on a desktop surface (see Figure 2.1). VideoDesk is an extension of VideoPlace (Krueger et al., 1985; Krueger, 1990), an early system that enables whole-body interaction with digital images projected onto a wall display. The VideoDesk system uses a video camera mounted above the desktop surface, image processing, and calculations about hand position to enable gesture-based interaction with digital materials projected onto the display.

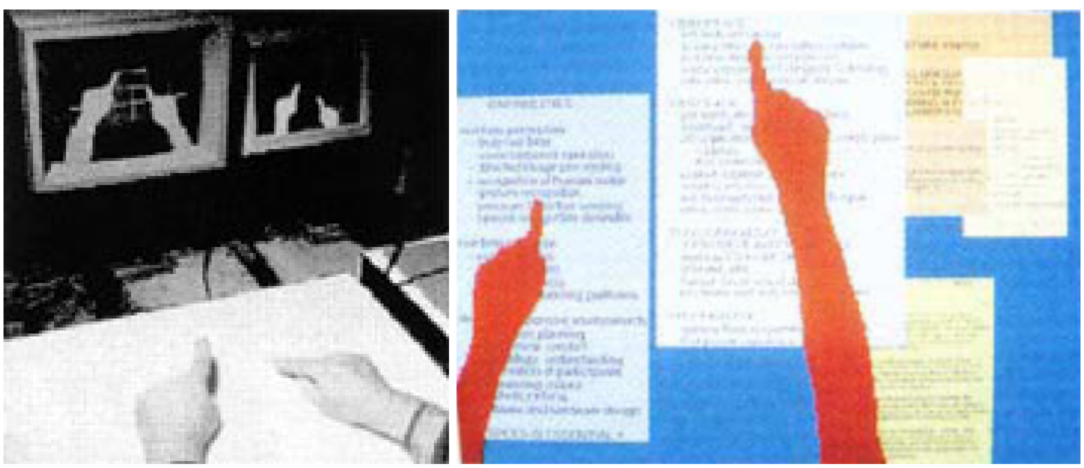


Figure 2.1: *Left:* Myron Krueger's VideoDesk, image borrowed from Krueger (1991). *Right:* two users in different locations point to documents projected onto the VideoDesk, image borrowed from Krueger (1993).

Pierre Wellner's Digital Desk (1991; 1993) explores augmented interaction on a desktop surface (see Figure 2.2). A projector mounted over the desk displays graphical information directly on top of items in the workspace, including paper documents. An overhead camera tracks hand and object movement, determining user interaction with the surface. This system is single-user, yet it demonstrates several fundamental ideas that continue to shape current interactive tabletop and desktop workspaces.

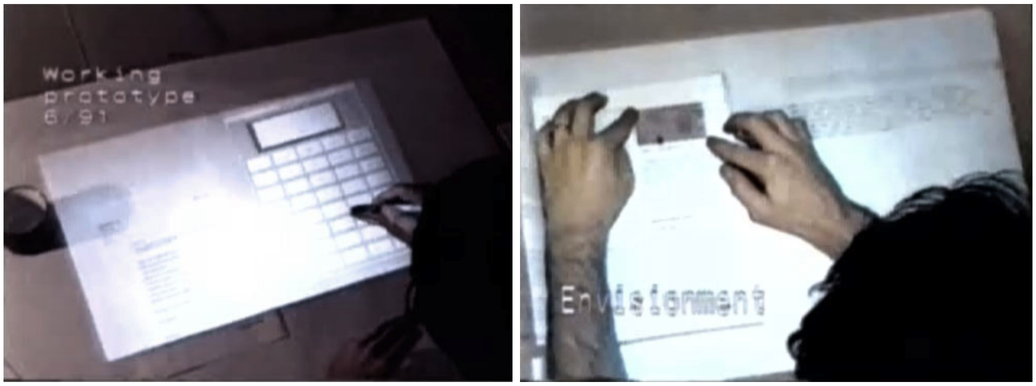


Figure 2.2: Pierre Wellner's Digital Desk integrates touch and pen input with paper and digital documents.

Another seminal project is Rekimoto's Augmented Surfaces (1999). This work demonstrates the potential for seamless multiuser interaction around the tabletop. Users can interact with digital objects projected onto multiple displays, including a tabletop, wall, and laptops (see Figure 2.3). Through a technique called Pick&Drop, users can move digital objects from one display to another. This system also integrates physical objects into the workspace, allowing users to drag information to and from paper-based materials.

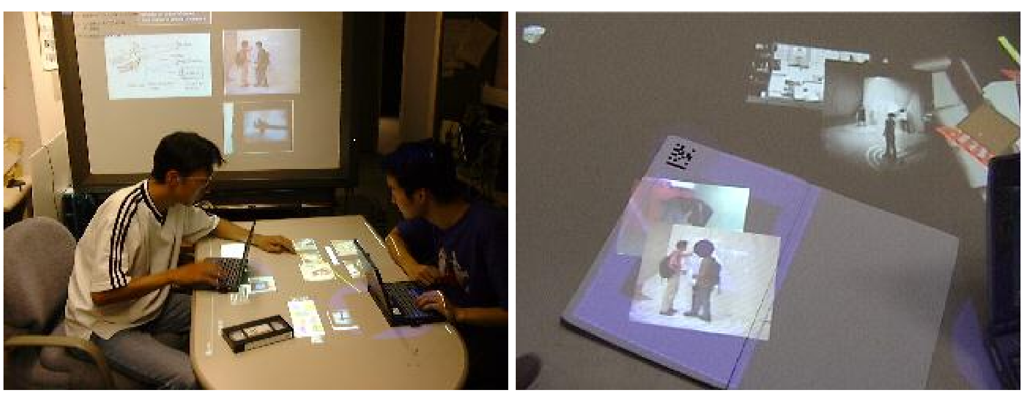


Figure 2.3: Jun Rekimoto's Augmented Surfaces supports fluid interaction between multiple devices and between digital and physical materials.

These systems showcase many of the techniques researchers in HCI are still exploring today. These early systems present important ideas, such as gesture-based interaction, dragging digital content from one device or workspace to another, fluid interaction between information in digital and paper form, and extend the notion of desktop computing to include a variety of devices and display configurations. The next sections describe more recent research on various interaction techniques present in these early systems.

2.2.2 Interaction Techniques

Augmented table and desktop workspaces exploit a variety of interaction techniques, including single-touch and multitouch, stylus and pen-based interaction, incorporation of mobile devices and laptops, physical objects or tangibles, and multimodal commands. I summarize these techniques below.

Touch Interaction. One of the primary modes of interaction with large horizontal interactive workspaces involves single or multiple-finger touch input. I focus on multitouch implementations in this section. Several techniques exist for instrumenting interactive workspaces with multitouch sensing. These include computer vision techniques involving frustrated total internal reflectance (FTIR), diffuse illumination, infrared lasers, and systems that exploit capacitive sensing or capacitive coupling.

FTIR implementations involve sensing from a camera under a surface illuminated by infrared light (Han, 2005). As a user touches the interactive surface, their finger disperses infrared light (see Figure 2.4, right). This light is detected by the infrared

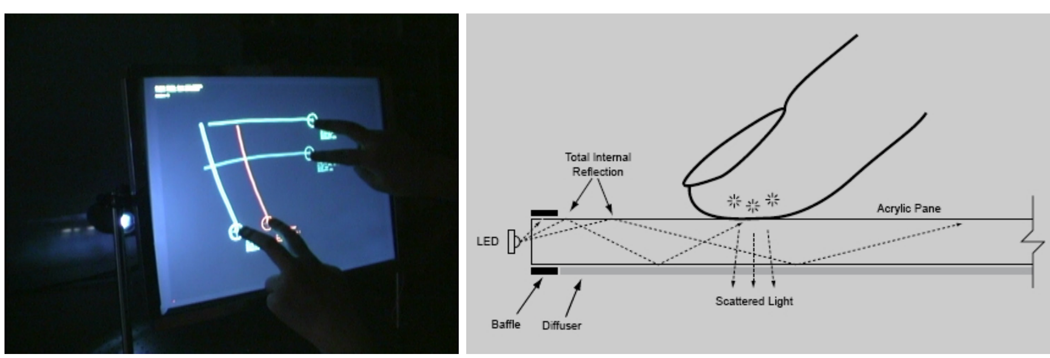


Figure 2.4: *Left:* Multitouch interaction with an FTIR display. *Right:* Schematic of FTIR sensing. Images borrowed from Han (2005).

camera mounted under the interactive surface. A variety of programming libraries facilitate processing of the camera images (e.g., Touchlib, 2011). Tabletop systems that use FTIR technology have two important advantages over other solutions; they are often low cost to create and can detect an arbitrary number of contact points of varying size and material. Adding additional cameras above the workspace enables sensing fingers and objects as they approach the interfaces (e.g., TouchLight, Wilson, 2004).

Many recent touch-based cell phone and tablet computers exploit capacitive sensing (e.g, Android phones; Apple iPad, iPod Touch, and iPhone). Capacitive sensing involves detecting the natural electrical charge carried by our bodies when we come into contact with an interactive surface. An application of this concept is the Mitsubishi DiamondTouch table (Dietz and Leigh, 2001, see figure 2.5). The DiamondTouch table is different from other capacitive systems because it relies on capacitive coupling. That is, with the DiamondTouch table, users sit on conductive pads that are connected to the tabletop display. A uniquely identifiable current is transmitted through the conductive pads and then through each person's body. The sensing surface is an array of antennas that recognize the unique signal passed through each person's body. Therefore, the

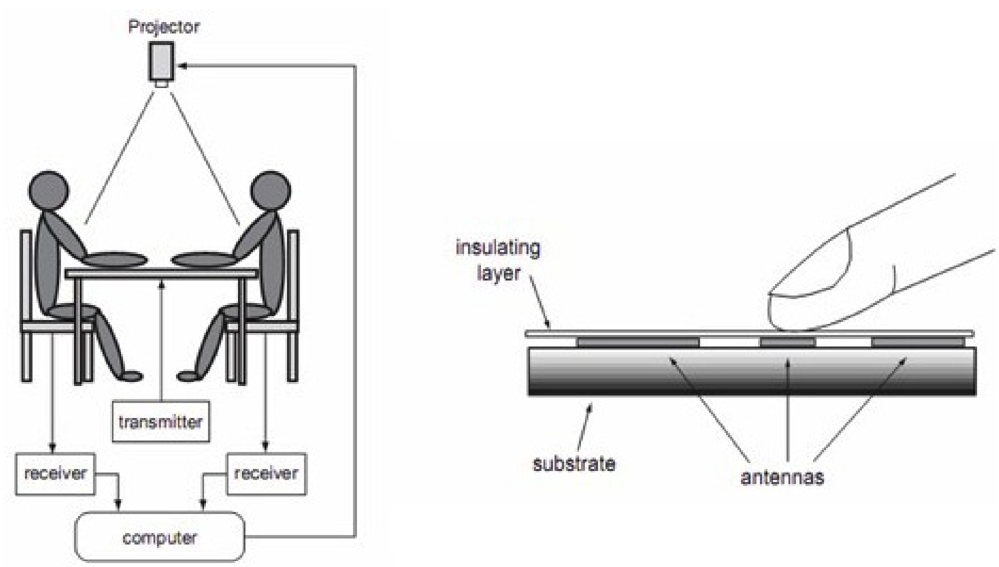


Figure 2.5: *Left:* Users are capacitively coupled to the DiamondTouch display. *Right:* Schematic of the DiamondTouch capacitive sensing surface. Images borrowed from Dietz and Leigh (2001).

system can distinguish which person is touching where on the interface, a challenging problem for computer vision based systems. There are two main limitations of the DiamondTouch: the system can only accommodate up to four users (each touching with two contact points), and users must remain capacitively coupled to the tabletop by sitting or standing on the conductive pads. Although the DiamondTouch has limitations for multitouch input, a number of multi-finger and whole hand gestural interaction techniques have been identified (Wu and Balakrishnan, 2003). Finally, the DiamondSpin toolkit (Shen et al., 2004) provides a Java API for rapid multitouch applications on the DiamondTouch table. In contrast to the DiamondTouch table, another system that uses capacitive sensing is SmartSkin (Rekimoto, 2002). This system can sense multiple hand positions and shapes as well as calculate how far hands and objects are from the surface, making it possible to track above-the-surface gestures.

Stylus and Pen-Based Input. Another common interaction technique for large horizontal workspaces involves the use of stylus and pen-based input. With a stylus, generally the underlying display technology senses interaction and performs necessary functions. ConnectTables (Tandler et al., 2001) is an example of a stylus-based interaction in a collocated workspace. This system involves tablet computers, and users are able to connect multiple tablets together to interact with other people and exchange content. Coeno is a project that allows multiple people to simultaneously create pen-based annotations on a tabletop display (Haller et al., 2005). A slightly different use of stylus input is demonstrated by the TractorBeam project (Parker et al., 2005). This implementation involves a tethered stylus and receiver that allows users to point at objects with the stylus. Their findings suggest that, compared to touch input alone, a stylus with the "tractor beam" technique is an appropriate and comfortable way to reach distant objects on a large table. A more recent example is work by Hinckley et al. (2010) involving multimodal commands that combine a stylus and touch input.

Computationally enhanced writing pens facilitate novel methods of interacting with digital documents on a computer display or projected workspace. Furthermore, digital pen technology transforms traditional paper documents into interactive digital artifacts. Compared to stylus-based interaction, where the display is sensing interaction, digital pens have embedded hardware that detects where the pen tip is located on a piece of paper and process that input.

Integrating paper documents into these collaborative digital workspaces is a powerful concept. It has been demonstrated in numerous settings that paper is a flexible medium with many affordances not supported by current digital systems (Sellen and Harper, 2001; Luff et al., 1992; MacKay, 1999; Nomura et al., 2006). A hybrid paper-digital approach is optimal for many contexts. Linking paper documents with a digital system has been investigated in numerous research projects. A few examples include the Digital Desk (Wellner, 1991, 1993), MacKay's A-Book (Mackay et al., 2002), and the Audio Notebook (Stifelman et al., 2001), and there are many more.

Two projects integrate digital pen technology with multitouch tables. The Diamond's Edge project (Jiang et al., 2007) combines an Anoto pen with the DiamondTouch table to support flexible sketching and sharing of ideas in public and private spaces. DigiPost (Bernstein et al., 2006) also couples an Anoto pen with the DiamondTouch table. This project links post-it notes with digital documents projected onto the tabletop display.

Mobile Devices and Laptops. Often it is useful to integrate devices such as laptops, cell phones, and personal digital assistants (PDAs) into a collaborative tabletop workspace. As mentioned earlier, a well-known example of integrating laptops with an interactive tabletop workspace is the Augmented Surfaces project (Rekimoto and Saitoh, 1999). This project lets users drag items from a laptop onto a digital tabletop surface and vice versa. Similarly, the UbiTable (Shen et al., 2003) allows two users to share content by dragging items to an area on their laptop that then sends the information to a shared tabletop display. The Caretta system (Sugimoto et al., 2004), designed to support urban planning tasks, provides users with a shared map on a tabletop display and private information on connected PDAs. Finally, the STARS gaming system (Magerkurth et al., 2004) allows players to interact through a shared tabletop display while receiving private information through a PDA.

Physical Objects and Tangibles. It is also possible to interact with a tabletop surface through physical objects, often called a tangible user interface (TUI). The design of these systems follows the hypothesis that physical materials have certain strengths that are not available in digital environments and vice versa. Two of the first examples of integrating physical objects with tabletop displays include the Urban Planning Table (Arias et al., 1997) and the metaDESK (Ullmer and Ishii, 1997). Both of these early systems were designed to understand the strengths and weakness of each media and propose new opportunities for human-computer interaction. Sensetable (Patten et al., 2001) advanced tangible tabletop interaction because the system could recognize physical objects more quickly and enable real-time interaction with the physical objects (e.g., rotating a dial).

A more recent approach involves SLAP widgets, or silicon illuminated active peripherals (Weiss et al., 2009). This technique combines the sensing of an FTIR table with translucent widgets such as a keyboard, dials, sliders, and buttons (see Figure 2.6). The main benefit of this approach is that users receive tactile feedback by interacting with tangible controls, but the system can update the status and display values of these control dynamically.

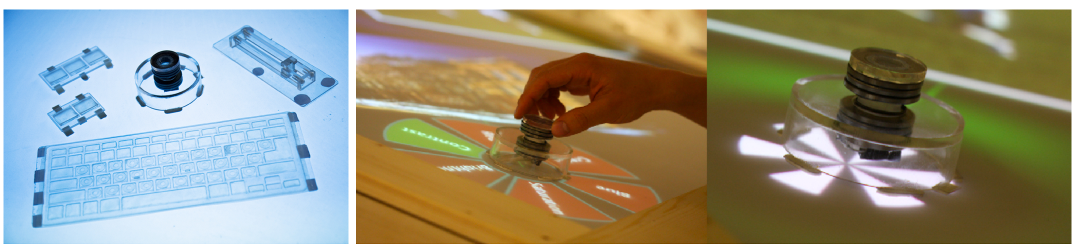


Figure 2.6: Slap widgets for an FTIR table. Images borrowed from Weiss et al. (2009).

Another influential system is the Lumisight Table (Matsushita et al., 2004). This system uses cameras and RFID technology to track finger position and tagged objects. Compared to previous systems, this implementation includes a rear projected image from four projectors.

One of the first systems to provide both visual and physical feedback to users is the Actuated Workbench (Pangaro et al., 2002). This system uses computer vision to track objects and magnets to move physical objects on the surface.

Multimodal Techniques Multimodal interaction with technology has been around since 1980. One of the first computer systems to incorporate multimodal input as commands is "Put-That-There" by Bolt (1980). The system tracked gestures and performed speech recognition as a way to interpret user commands. More recently there have been a number of multimodal systems for horizontal workspaces designed to support military command and control scenarios. Quickset (Cohen et al., 1997) is a tablet computer system that allowed pen and voice interaction with a map. Similarly, their subsequent multimodal systems involved speech and pen-based interaction with paper media (e.g., Cohen et al., 2008). Corradini et al. (2002) created a system that incorporates speech and three-dimensional gestures over maps displayed on a drafting board. Tse's dissertation research (2007) includes several important contributions to this area. Specifically, Tse and his colleagues developed a wrapper that transforms single-user applications into multiuser, multimodal tabletop applications (Tse et al., 2006), studied how pairs interact around a multimodal tabletop display (Tse et al., 2007b), and explored multimodal tabletop gaming scenarios (Tse et al., 2007a, see Figure 2.7, left). Finally, the Reflect project (Bachour et al., 2008) used a traditional table embedded with LEDs to display attributes of conversation by participants, such as the percentage of time spent talking (see Figure 2.7, right). The system works by localizing speech through a three-microphone system and an audio source separation algorithm.

Figure 2.7: *Left:* Multimodal tabletop interaction with World of Warcraft game, image borrowed from Tse et al. (2007a). *Right:* Reflect tabletop conversation system, image borrowed from Bachour et al. (2008).

2.3 Applying Distributed Cognition

As illustrated in the previous section, the field of human-computer interaction is experiencing a shift from single-user computer systems that support isolated activity to multiuser hybrid technology workspaces that incorporate a range of digital and physical media. Just as the transition from command line entry to graphical user interfaces (GUIs) radically changed interaction with computers, novel gesture-based, whole body, and multimodal interaction techniques are reshaping how we interact with digital information and other people. Distributed cognition provides an ideal theoretical framework for understanding human interaction in digital workspaces, informing technology design, and evaluating the impact of new digital tools on interaction (Hutchins, 1995; Hollan et al., 2000; Hollan and Hutchins, 2009; Rogers and Ellis, 1994; Halverson, 2002)

Unit of Analysis. The theoretical framework of distributed cognition (Hutchins, 1995; Norman, 1993; Salomon, 1993) is the foundation for how I understand human interaction in computer-mediated workspaces. Distributed cognition is closely related to *activity theory* (Leont'ev, 1978; Vygotsky, 1978; Wertsch, 1981; Cole, 1985; Engeström, 1990). See Nardi (1996) for one discussion of the key differences and similarities between these theories with respect to HCI research. Importantly, in a distributed cognition approach, the unit of analysis is the larger system involved in a cognitive activity, including multiple people, their actions, artifacts in the environment, cultural practices, and the accumulation of these representations and processes over time. Furthermore, what is included in the cognitive system changes depending on the activity or sub-activity of interest. In my research, the cognitive task is effective communication, or reaching a shared understanding about a topic or problem in a shared face-to-face conversational space. The unit of analysis often includes multiple people (e.g., a therapist and client, or a doctor and patient), the digital and physical artifacts involved in interaction, the immediate context in which activity occurs, and the cultural practices each person brings with them about how communication is accomplished. One aspect of distributed cognition that is particularly powerful for this thesis is the notion that our current cognitive activity is shaped by our past activities, and this current activity also shapes our future interactions. This notion can be taken quite literally using several mechanisms I introduce in this thesis to capture, represent, and archive spoken dialogue in a face-to-face conversation.

Embodiment. Another aspect of distributed cognition that makes it a productive framework for understanding interaction with multimodal technologies, particularly gesture-based systems, is its treatment of the human body and our actions as central to cognition. Cognition is an emergent property of interaction, and as such the human body plays a central role in our cognitive activity (Hutchins, 1995). Theories of *embodiment* (Clark, 1997; Johnson, 1987; Nunez, 1999; Varela et al., 1991) provide additional support for understanding how our physical bodies both aid and constrain how we interact with and reason about phenomena in the world. These theories come at an exciting time for HCI research given the critical need for deeper understandings of gesture-based and whole body interaction with computer systems (i.e., enabled by new technologies such as Microsoft Kinect[1]).

Method and Analysis. A distributed cognition approach emphasizes ethnographic methods, including *in situ* observations, detailed field notes, interviews, and artifact collection and analysis. The technique of *cognitive ethnography* combines traditional ethnographic methods with micro-analysis of cognitive systems in action (Hutchins, 1995). In some situations, is it not possible to conduct ethnographic work. When ethnographic observations are not immediately possible (e.g., the case of my research on medical communication), it is productive to conduct laboratory experiments to examine interaction. However, distributed cognition allows us to look differently at these experiments. Hollan et al. (2000) state:

> An experiment is, after all, just another socially organized context for cognitive performance. This means not only that we look at so-called *real-world* settings, but that we look differently at experiments, seeing them as settings in which people make use of a variety of material and social resources in order to produce socially acceptable behavior.

Whether data is collected in the field or laboratory, descriptive analysis at a micro-level is a common technique for understanding and communicating cognitive processes of interest. Micro-analysis of human interaction involves understanding patterns and the organization of speech, gesture, eye gaze, and the body similar to that of ethnomethodology (Garfinkel, 1967; Schegloff and Sacks, 1973; Schegloff, 1984;

[1]http://www.xbox.com/en-US/kinect

Goffman, 1981; Kendon, 1967, 1980; Goodwin, 1981, 1986). An important aspect of interaction analysis is that no single modality (e.g., speech) is understood fully without examining accompanying modalities of interaction, the context of activity, and the material workspace. That is, it is important to consider a segment of interaction as a situated activity system. For example, within the context of airport personnel looking at airplanes, Goodwin and Goodwin (1996) describe the act of seeing as a situated activity. Hutchins and Palen (1997) examine multilayered representations produced within the cockpit of a commercial aircraft, exploring the critically interconnected nature of speech, gesture, and artifacts in space that together create a coherent interaction. Goodwin (2003b) analyzes the situated nature of pointing acts, illustrating that meaning is constructed through the juxtaposition of multiple semiotic fields. Furthermore, these modalities are not merely additive; they build upon one another through complex relationships among a range of resources, central to which is the surrounding environment impinging on activity (Hutchins and Palen, 1997; Alač and Hutchins, 2004).

2.4 Traditional Tables and Physical Media

Prior research has examined work patterns around traditional tables and desks to inform the design of horizontal interactive workspaces. Examining how people interact with physical media also provides inspiration for interactions with digital media.

Studies of personal space, also known as proxemics, can further expand our understanding of interaction around traditional horizontal workspaces (Hall, 1966). Tang (1991) made several important observations about small group activities around traditional tables. In particular, he noted that people use hand gestures on and around the tabletop to communicate information. His motivation for studying small group activity around traditional tables was to inform the design of subsequent collaborative technology. In terms of work flow around tables, Tang (1991) observed that people tend to divide the table surface into separate work areas. Groups treat the center of the table as public space and outside edges as private areas (Scott et al., 2004). Tabletop activity typically involves transitions between closely coupled actions with another person and individual actions (Elwart-Keys et al., 1990; Mandviwalla and Olfman, 1994).

Examining interaction with physical objects on a table also informs the way we design interactions with virtual objects on digital tables. The orientation of objects on a shared work surface affects the legibility of text-based content. Kruger et al. (2003) found that the orientation of objects on tables also impacts coordination and communication among group members. Researchers have studied activity around traditional board games to understand how orientation of objects affects comprehension (e.g., Whalen, 2003). Observations of these problems on physical tables led to techniques for reorienting objects on digital tables (see Ringel et al., 2004a).

Comparing interaction with physical and digital media on tables can inspire the design of novel interaction techniques. Work by Terrenghi et al. (2007) highlights two key affordances of paper media compared to digital media on a surface computer. First, paper allows manipulation of objects in three-dimensional space whereas objects on a tabletop computer are presented and accessed in a two-dimensional form. Second, paper provides multimodal feedback about interaction through touch in a way that touching an interactive surface to manipulate digital objects does not. This research on interaction around traditional tables and interaction with physical objects illustrates the need to develop hybrid workspaces that integrate the digital and physical world.

2.5 Understanding & Designing Interactive Workspaces

Increasingly, researchers in HCI are studying the complexities of human interaction in technology-rich workspaces. Research within computer-supported cooperative work (CSCW) often applies ethnomethodology to the study of sociotechnical systems. For example, Heath and Luff (1991, 1992) study cooperative activity and the coordination of attention in London underground control rooms. Hindmarsh and Heath (2000) examine how people establish mutual orientation towards objects of interest, including documents and computer displays. Similarly, Luff and Heath (2000) characterize the nature of collaboratively produced commands to computer systems, noting that these commands are created moment by moment among workers in the space and that the division of labor occurs over the course of the interaction.

There is a growing interest in gesture and its relevance to the design of cooperative computing systems. Tang (1991) noted the pervasive nature of hand gestures in a group drawing activity and indicates the need to better understand this activity in relation to the activity and artifacts in a collocated workspace. Bekker et al. (1995) studied gestures found in face-to-face meetings to examine the role of gesture in face-to-face design activities as a way of informing the design of cooperative systems. Kraut et al. (2003) examined how visual information, deictic reference in particular, enabled situational awareness and conversational grounding in face-to-face, video-based, and audio-based interaction.

Research also examines the benefits of a horizontal computer workspace compared to a vertical display. Rogers and Lindley (2004) noted an increase in the use of gesture when groups interacted around a tabletop display compared to a vertical whiteboard display (the details of various gesture forms observed are not provided, though it appears that they are primarily referring to deixis). They also found that tables enable a face-to-face work style, which promotes more frequent turn-taking and more equitable participation than a vertical whiteboard setup (Rogers and Lindley, 2004). The shoulder-to-shoulder working style around a whiteboard results in a single leader controlling interaction, whereas the face-to-face setup of a tabletop allows frequent role switching and shared leadership responsibilities. In another study, Rogers et al. (2004) found that touching a display with fingers has ancillary benefits for group work, such as supporting turn-taking and as a way to emphasize or function in place of speech acts. With respect to gesture, Tse et al. (2007b) provided similar observations of pairs interacting around a multimodal tabletop display. They note that “speech and gesture commands serve double duty as both commands to the computer, and as implicit communication to others,” (Tse et al., 2007b).

In my research I apply these perspectives on interaction around horizontal workspaces to the design and evaluation of novel systems that augment face-to-face communication. It is worth elaborating one key example of this approach that informs research within this thesis: the eSPACE project (Rodden et al., 2003; Halloran et al., 2003; Rogers and Rodden, 2003). This project begins with a six-month field study on travel agents interacting with clients around their current desktop workspaces. Field research indicates an unbalanced experience between the two participants. The travel agent has privileged access to information displays, leaving the client to piece together information to make decisions about their travel. Based on this understanding, the research team redesigned the workspace to better support face-to-face interaction between collocated participants. Figure 2.8 presents the original and redesigned workspaces.

Figure 2.8: *Top:* Current workspace setup and face-to-face consultation model for travel services. *Bottom:* Redesigned workspace with improved interaction model. Images borrowed from Rodden et al. (2003).

2.6 Conclusion

This chapter presented technical advances in the design and implementation of multiuser interactive workspaces. I examined technology development in light of theories of distributed cognition and sociotechnical analyses of existing technology systems. Through this I highlighted the need for a deeper understanding of human interaction throughout the design and evaluation of new computationally enhanced workspaces, particularly those which focus on augmenting communication.

Chapter 3

SIDES

This chapter presents a design case study of SIDES: Shared Interfaces to Develop Effective Social Skills. SIDES is a tool designed to help adolescents with Asperger's Syndrome practice effective group work skills using a four-player cooperative computer game that runs on tabletop technology. We present the design process and evaluation of SIDES conducted over six months with a middle school social group therapy class. Our findings indicate that cooperative tabletop computer games are a motivating and supportive tool for facilitating effective group work among our target population. This project reveals several design lessons that can guide the development of similar systems.

3.1 Introduction

Interactive table technologies are a new medium through which adolescents who have difficultly learning to work effectively in group situations can practice group work skills in a supportive and motivating way. Tabletop technology encourages group interaction around one interface in a way that other computer workstations and video gaming systems do not. Computationally-enhanced tables allow face-to-face interaction and multiple simultaneous inputs from a group of users. Applications designed to run on tabletop technology can require user specific actions and cooperative actions (Morris et al., 2006). We implemented SIDES, a four-player cooperative computer game for social group therapy, on the DiamondTouch table (Dietz and Leigh, 2001), an interactive tabletop surface that can receive multiple simultaneous inputs and uniquely distinguish each user's touch. This functionality allows application designers to restrict or require input from certain users during the tabletop activity. SIDES leverages this aspect of tabletop technology to encourage cooperative decision making and equitable participation by group members. These aspects of group work are particularly difficult for adolescents with Asperger's Syndrome to learn and for their social skills therapists to moderate in traditional group work situations.

Asperger's Syndrome (AS) is a Pervasive Developmental Disorder and is considered an Autism Spectrum Disorder. Statistical data on the prevalence of AS is unclear, as many cases go undiagnosed or are misdiagnosed. It is estimated that AS occurs in 3.6 to 7.1 of 1000 children (Ehlers and Gillberg, 1993). While children and adolescents with AS often do not have significant delays in cognitive and language development, these individuals have difficulty understanding accepted social conventions, reading facial expressions, interpreting body language, and understanding social protocols. These social deficits can lead to challenges in learning effective group work skills, including negotiation, perspective taking, active listening, and use of pragmatic language.

Adolescents with AS often describe the computer as a comfortable and motivating medium. Through our approach we leverage the comfort of working with a computer to help these individuals practice effective group work skills. Our evaluation of SIDES reveals benefits inherent in the use of tabletop technology as a therapy tool for this audience and discusses the tradeoffs of supporting this type of group activity with computer- versus human-enforced rules. Through our analysis, we found that the affordances of tabletop technology – specifically the ability of tables to facilitate face-to-face interaction, allow simultaneous user input while controlling individual user actions, and encourage cooperative decision making – were critical in successfully supporting cooperative work with our target user group and have valuable implications for the broader CSCW community.

3.2 Background

The majority of computer programs for social skills development for our target audience are designed for one user working directly with the application and lack the face-to-face interaction found in authentic social situations (Baron-Cohen, 2004; Team Asperger, 2004). Social skills therapy groups help adolescents with AS learn strategies to navigate social situations. Social skills therapists who lead these groups often use card and board games to help adolescents practice appropriate social interaction techniques with peers. These traditional games, however, may not sustain interest or motivate students enough to overcome challenges in social interaction. Traditional board games can be inflexible and may be difficult to modify to support current classroom topics and learning goals.

On the other hand, tabletop technology is a unique platform for multi-player gaming that combines the benefits of computer games with the affordance of face-to-face interaction. Tabletop computer games have been explored for general audiences (e.g., Magerkurth et al., 2004; Mandryk and Maranan, 2002), but prior to this project had not been designed for a special needs population who would benefit from social computer games.

3.3 Related Work

There are currently a number of single-user computer programs to help with social skills development. These existing applications typically focus on rote memorization of facial expressions and emotions. Examples include Mind Reading: The Interactive Guide to Emotions (Baron-Cohen, 2004) and Gaining Face (Team Asperger, 2004). Memorization of social cues may be helpful to some adolescents, but this isolated activity lacks a supportive and authentic context for application of these skills. Teaching appropriate social protocols with virtual reality has also been explored (Beardon et al., 2001; Kerr et al., 2002). Despite advances in facial imaging, it is difficult for computers to completely replicate the nuances of human social behavior. Though social cue memorization and virtual reality applications are valuable, neither of these approaches provides a fully supportive and authentic means of practicing effective group work skills.

The goal of our application is not to teach skills explicitly, but rather to provide a motivating and supportive experience through which adolescents can practice social and group work skills discussed in group therapy sessions. The pedagogical design of

SIDES stems from Vygotsky's theory that learning is a social process and has its roots in social interaction (Vygotsky, 1978). Collaborative activities and cooperative games have been shown to specifically benefit individuals with AS (Kerr et al., 2002). Video games have been used to facilitate therapy and rehabilitation for certain special needs audiences (Griffiths, 1997), but not as a cooperative activity for this audience. SIDES leverages these educational theories and the prior work described above to provide an authentic and engaging activity to supplement current therapy techniques for teaching social and group work skills.

The term *single display groupware* (SDG) refers to systems that support co-located, computer-supported cooperative activity around a single, shared display (Stewart et al., 1999). Interactive tables, such as the DiamondTouch table (Dietz and Leigh, 2001), are a form of SDG that promote face-to-face interaction. The horizontal setup is in contrast to the shoulder-to-shoulder style of interaction promoted by vertical, wall-mounted displays (Rogers and Lindley, 2004). Furthermore, individualized control over input devices with SDG has been linked to increased performance and achievement in computer games with adolescent pairs (Inkpen et al., 1995). Tabletop technology has the functionality to provide each group member with individualized control over the interface, which may be particularly useful for adolescents with AS who describe the control they have over computers as comforting.

3.4 Design Process

We conducted observations, interviews, and paper and digital prototype tests over a period of six months with middle school students (11-14 years old) and therapists from a social skills therapy group. Twelve students and their school-designated social skills therapist were involved in this study. While the majority of students in our study have a primary diagnosis of AS, other students from this class who participated in the study have social skills challenges stemming from other disorders, including diagnoses of High-Functioning Autism, Attention Deficit Hyperactivity Disorder (ADHD), Apraxia, and Klinefelter's Syndrome. Our methodology for understanding the needs and learning goals of this population included participant observation as well as group and individual interviews. We employed a participatory design approach (Schuler and Namioka, 1993), as this inclusive design process is described as essential and critical in creating a satisfactory solution when designing for special needs populations (Fischer and Sullivan, 2002). Our research team involved students and adults with AS, social skills therapists, and parents of children with AS in all aspects of design and evaluation.

3.4.1 Design Goals

Our goal was to develop a cooperative, multi-player tabletop computer game that encourages the meaningful application of group work skills, such as negotiation, turn-taking, active listening, and perspective-taking, for students in social group therapy. We intentionally designed SIDES to leverage the cognitive strengths and interests of individuals with AS. In our interviews with children and adults with AS, many individuals described an interest in visual games such as puzzles; as a result, we created a puzzle-style game. AS occurs in only one female for every four males (Ehlers and Gillberg, 1993), so we chose a game theme of frogs and insects in order to appeal to our predominately male, adolescent audience. For adolescents with AS, the challenge in playing any game is learning to work cooperatively and play fairly with each other; our goal with SIDES was to design a game that facilitates cooperative game play in a meaningful way.

3.4.2 Field Studies and Observations

As participant-observers of a middle school social skills therapy class, we sat with students and participated in group discussion on topics such as listening, turn-taking, and leadership. We attended seven sessions, each lasting approximately one hour, to investigate current approaches to teaching social skills as well as student interests in and out of the classroom. We conducted six interviews with middle school social skills therapists and speech pathologists to understand current teaching methods and classroom techniques and to identify potential solutions for teaching group work skills. Working with this user group was challenging, as illustrated by a comment from the social skills therapist who leads this social therapy group:

> Some of my kids go into mainstream classes and they just can't work with other people. We have to find the right mainstream kids who will have the patience and tolerance to deal with our kids' behaviors. Then some of our kids just flat out refuse to work in groups because they don't want to give up their power and control. Control for these kids is not something they have a lot of so they try to control their environment.

When we were able to conduct interviews with students from this class, we faced challenges in building rapport and making students feel comfortable during the interview. One student, for example, "shut down" during her interview. She would not make eye contact and only provided one-word answers to open-ended discussion prompts. Instead of one-on-one interviews, we adjusted our method of data collection and found that group interviews with four or five students from the class were more productive. Interviews with students from this class revealed discontent with current group therapy activities such as discussing emotions and reporting on weekend activities. We found that "game day" (therapy sessions where students play board games) was one of the few interview topics that elicited positive and excited responses from students. One seventh grade girl from this social group therapy class pointed out that the challenge in designing a motivating and exciting game is to avoid creating a game that appears overtly educational. This student is an avid gamer and is currently designing her own computer game. When asked how she would design a game to teach the social skills topics addressed in group therapy, she replied, "I don't know. I don't really like those types of games. I don't do educational games." She then explained that "entertainment games are just when you're doing them for fun" and educational games "start teaching you stuff and they get away from all the entertainment and fun." We realized that the challenge in designing a compelling cooperative game for this audience would be to create an engaging experience that does not directly focus on traditional content from social skills therapy sessions.

Games are a prominent theme that emerged from our observations and interviews. Students in this class frequently play online games and video games at home. We found that board games are often used as a tool during therapy sessions. Though classroom game time is particularly beneficial for this audience, the activity must be closely monitored to prevent verbal and physical altercations. The students' social skills therapist commented, “With these kids we have to be on alert when they are playing board games in class. We walk over at the first sound of voices raised. Other kids would be fine and could work out a disagreement, but with our kids we have to monitor behavior very closely and know when it’s time to intervene.” We realized that regardless of our game design, an adult may have to monitor game play for behavioral purposes.

3.4.3 Game Rules

We created a highly visual, four-player puzzle game and designed the rules to increase collaboration and decrease competition. At the beginning of a round, each player receives nine square tiles with arrows (three copies each of three arrow types). Unique arrow types (e.g., left, right, turn) are distributed among participants so that no participant has all 12 arrow types in their “hand.” Students are asked to work together to build a path with their pieces to allow a “frog” to travel from the start lily pad to the finish lily pad. There is a limited supply of each arrow type, thus encouraging students to cooperatively build an optimal path to win the most points. To gain points, the path must intersect with insect game pieces on the board. The insects are worth various point values (e.g., each dragonfly is worth 20 points). The group of students must agree on one path that collects the most points with their given amount of resources. Once all players agree with the solution, the frog will travel along the path and collect points by eating all the insects it encounters.

3.4.4 Paper Prototype

We tested a paper prototype of SIDES (see Figure 3.1) to finalize the rules, check for game balance, and determine whether this initial prototype showed enough promise to be turned into a digital game. The paper version of SIDES is ideally suited for four players, but more people can play with minor adjustment. We tested the prototype with two five-student groups from the social skills therapy class. After playing multiple rounds, we held a group interview and brainstorming session about the gaming experience. The students were positive about the game design and flow of game play. Students gave positive feedback on the frog and insect theme and offered numerous thematic suggestions. After observing both groups play the prototype, the students' therapist commented, "I was impressed with how they all shared the responsibility and actually played collaboratively rather than one person dominating... even those who are normally the least active in the groups were active and engaged the entire time." The paper version was successful in that it provided proof of concept for a cooperative game design. However, there are still significant advantages of a computer version for these adolescents. A computer game can enforce rules without the therapist having to police game play, thus freeing up his/her time to attend to higher level group work issues. Also, adolescents with AS typically find comfort in controlled and structured interactions with a computer, thus making a computer version even more promising.

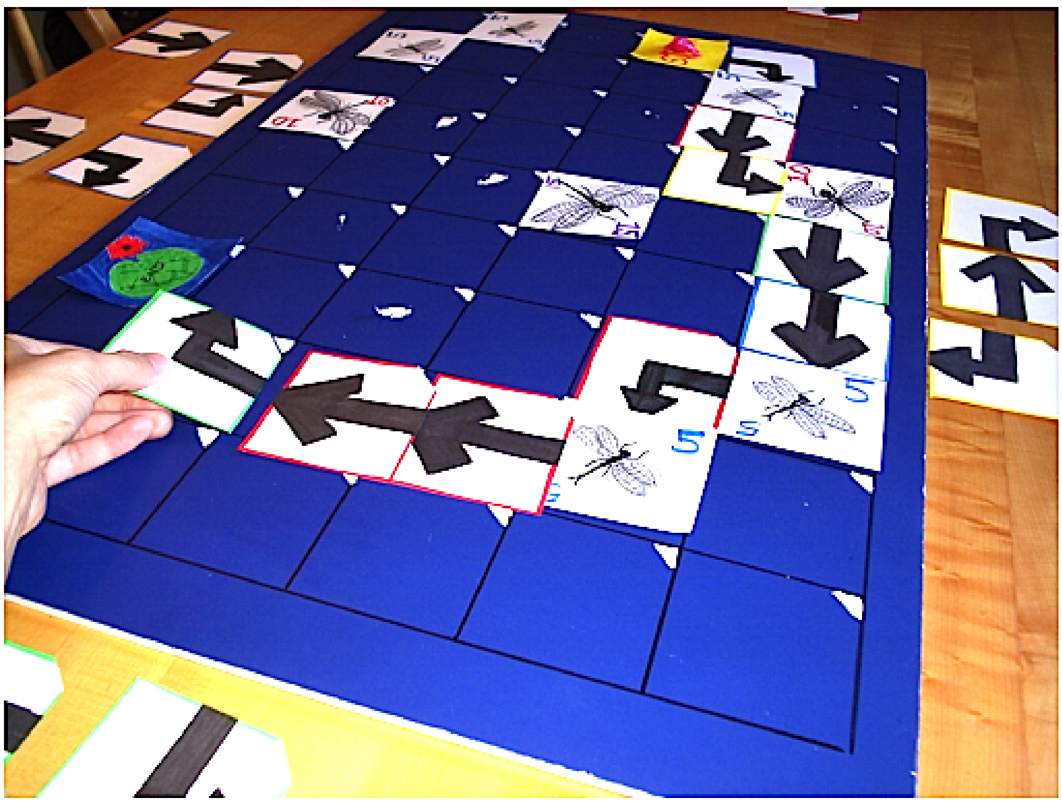


Figure 3.1: Paper prototype of SIDES game tested with middle school social skills therapy class.

3.4.5 DiamondTouch Implementation

After successful testing with the paper prototype, we implemented a computer version of the game in Java for the DiamondTouch table (Dietz and Leigh, 2001), a multi-user touch sensitive tabletop with a top-projected display. We wrote our application using the DiamondSpin tabletop user interface toolkit (Shen et al., 2004). As with the paper version, players seated around the table receive game pieces to place on the board and create an optimal path from start to finish. Game pieces with different types of arrows are divided amongst players and initially placed in piles directly in front of each person (see Figure 3.2, A). We chose this distributed configuration of game pieces based on findings from (Fischer and Sullivan, 2002), where the center area of the table

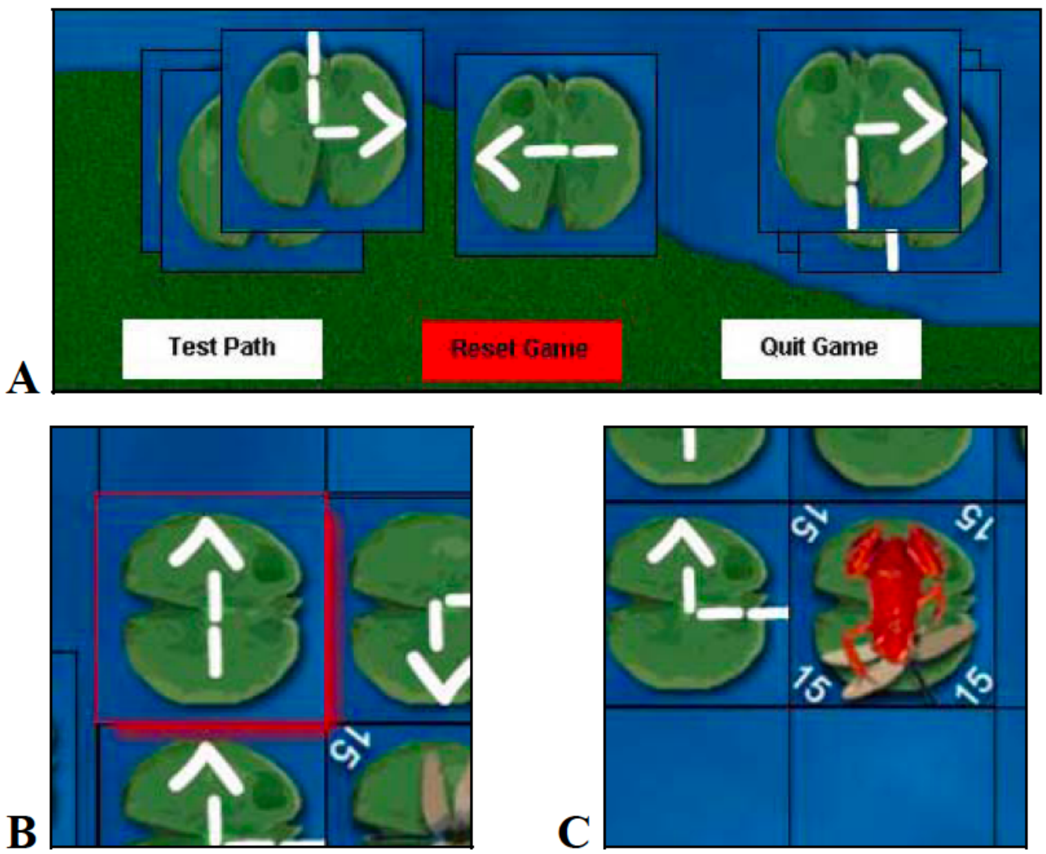


Figure 3.2: Interface components: A) Each player has a control panel with voting buttons located along the border of the table nearest each user's seat. B) Arrow pieces highlight with the player's color when touched. C) The frog "hops" along the path and eats insects to win points.

is perceived as a group space and areas directly in front of each person are considered spaces for personal items. We did not incorporate a timer or impose any time limits on the game to prevent students from feeling rushed and forgoing collaboration just to reach a solution. The computer version gives each player a control panel in the region of the interface closest to his or her chair. In each player's control panel are round and point indicators as well as voting buttons to test the path, reset, or quit the game (Figure 3.2, A). The voting buttons force the group to "vote" unanimously in order to change the state of the game. For instance, players must vote unanimously to test their path once a solution is reached by each activating their own "Test Path" button. This feature was implemented to ensure that no one player had more control over the state of the game than another player, and to encourage social interaction by necessitating communication and coordination with other members of the group. This first version of the computer game did not enforce rules, such as turn taking or piece ownership. This design decision was made so that the game remained more open-ended and we could investigate the minimal amount of structure necessary for encouraging effective group work.

3.5 Evaluation: Session 1

The primary research questions that guided the first evaluation of SIDES include:

- Are tabletop computer games an appropriate and feasible tool for facilitating social skills development for this audience?
- Do any sensory or motor issues specific to this audience affect interaction with tabletop technology?

3.5.1 Method

We tested our initial design with five students from the same social skills therapy class we observed and with whom we tested the paper prototype. These students were all male (mean age of 12.6 years, stdev=0.89) and in the same social skills therapy class. This group consisted of three adolescent boys clinically diagnosed with AS, one with Apraxia, and one with Klinefelter's Syndrome.

The DiamondTouch table is difficult to transport to a testing site, so students came to our university lab to play the digital version of SIDES. Several of the students' parents and their social skills therapist from school came to the lab at our university to oversee the testing session. Several parents requested that they be present in order for their son to participate, so they quietly observed the session from the other side of the lab. In more authentic, non-testing situations, we expect that parents would be present while their children were using SIDES, so the presence of parents during the testing session should not confound our results.

In this version of SIDES, the computer did not enforce rules. The therapist facilitated game play, monitored student behavior, and led a group discussion after play ended. Students played for two half-hour blocks of time. Following each half-hour play session, students discussed their experience with the therapist. The game is ideally suited for four players, so students rotated in and out after each round of play (where a round entails the successful completion of a path connecting the start and finish lily pads). This group played a total of six rounds, so each student averaged four or five rounds of play. Before beginning, the students were given a brief tutorial on how to use the DiamondTouch table and then instructed to work together to come up with one solution while playing SIDES. Two researchers observed the testing sessions and took detailed notes. The game play and discussion were videotaped for later analysis. All interactions with the interface were logged by the computer. Students individually completed a questionnaire after playing SIDES but before the group discussion.

3.5.2 Findings

We found that students remained engaged in the activity the entire time and were excited by the novelty of the technology. However, the students' excitement around playing a computer game on new technology in a new environment provided additional behavioral challenges. The students' therapist commented, "Even though their behavior was very positive, they were still talking over each other and not taking turns like we discuss in group therapy... they were really enthusiastic and had difficulty navigating back-and-forth conversation."

Figure 3.3: Four students from the social skills therapy class play SIDES during play testing Session 1.

Some students exhibited a high level of control over their behavior and made positive contributions to the group without dominating the activity. Drew, a seventh grader with AS, suggested several strategic moves to the group but was repeatedly ignored. Later he commented on the group's final solution, "It's not exactly like my planned route, but it's close enough." Drew's comment illustrates perspective taking, realizing that other people have different yet valid ideas, a topic that is frequently discussed in group therapy. Drew's mother observed the testing session from the other side of the lab. After the session she explained:

> I've actually found it rather interesting watching my son because he tends to be decisive about things and be more of a leader, but he's not forcing his will on anyone else here at all. He's listening and seemingly much more socially conscious than I think of him in terms of trying to be involved, but not trying to take over or get angry. So I'm actually quite pleased to see that.

In contrast, some non-cooperative behaviors indicate that additional structure could have helped other adolescents control their impulse to dominate the activity. Several rounds of play were chaotic with kids pushing each others' hands off the interface and yelling loudly. One outspoken student often took control of the game, reaching across the table to move other players' pieces without asking and telling others which piece to play next without eliciting input. This student's father observed the testing session and commented:

> With [my son], tact and making other people feel good about what they're doing doesn't even enter the equation... he'll try to get the ideal result of whatever problem is in front of him and how that impacts other people doesn't even occur to him. That's what he needs to learn more of. Games like this give him more practice.

In the debrief immediately following the gaming session, the students gave an overwhelming response regarding the need for order while playing. One commented, "There always has to be a leader; otherwise it will be wild and nobody will get anything from it." In response to this comment, Brad, a seventh grade student, stated, "We're supposed to work together. We're supposed to be equals." Brad was the quietest participant during the testing session and quickly became agitated and covered his ears when his peers spoke loudly at each other. During an individual follow-up conversation with Brad, he explained, "Last time it was chaos." He looked at the ground and paced back and forth, "yeah, it was really chaotic until I got to be the leader." By "leader" Brad is referring to a point in the session where the therapist closely monitored the students and gave each a chance to make decisions for the group.

In this first round of testing, we wanted to assess the appropriateness of tabletop technology for this audience. Our primary concern was whether these adolescents could learn sufficient control over the interface given the tactile input required by tabletop surfaces. Participants answered "How hard was it to move the pieces around on the table?" with a mean of 2.2 (stdev=0.45) on a five point Likert scale (1 = "not at all difficult" and 5 = "extremely difficult"). Based on the self-reported response by students and observations by the two researchers running the session, we conclude that participants found the mechanics of using the touch-sensitive tabletop technology manageable.

Overall, the students found SIDES to be a highly motivating yet challenging experience. After playing, one eighth grade student remarked, "Are we going to play again? I want to play it in the classroom." According to the students' therapist, this excitement carried over into the classroom for several days after the session and caused group discussions about the gaming experience, allowing him to tie the experience back into current classroom social skills topics. Session 1 demonstrated the promise of tabletop computer games as a tool for helping our target user group practice effective group work skills, as the students were engaged and motivated to participate.

3.6 Design Iteration

Play Testing Session 1 revealed that SIDES was motivating for these adolescents and is a promising tool for supporting effective group work among this user population. Session 1 also suggested that explicit game rules, such as turn taking and piece ownership might help reduce controlling behaviors of some students and encourage less engaged members to be more involved in the activity. We revised the game to include computer-enforced turn taking and restricted access to game pieces, as per our observations and feedback from the students' therapist. After Session 1, the therapist suggested, "Whoever's turn it is should be the only one who can manipulate the pieces. You can see that the kids can't keep their hands off. They will reach over and if one kid is too slow or taking in more information, they might not be able to wait and will break the rules by stealing another person's piece." The computer provides hard, fast, and consistent rules in a way that the therapist as a human facilitator cannot. The rule enforcement was enabled by the DiamondTouch table's ability to distinguish between four distinct users and to associate a user identity with each touch input. Thus, the "piece ownership" rule only allowed players to move the pieces that were given to them at the beginning of the game; attempts to move other users' pieces were ignored by the system.

We also redesigned the control panel in front of each player to include a turn taking button (Figure 3.4). Each player's turn taking button indicates whether or not it is that player's turn. A player may make as many moves with their own pieces during their turn as they like. The player whose turn it is has control over when they end their turn by pressing their turn taking button. We implemented a “give” protocol as described by (Inkpen et al., 1995, 1997) to prevent one student from “stealing” control from another player. Our intent behind implementing this turn-taking mechanism was to give players a feeling of ownership over the activity while still encouraging negotiation between players. Our rationale for not implementing a game timer was to prevent players from feeling rushed or pressured to think quickly and make a move, as this may lead to anxiety, disruptive behavior, and discourage cooperative discourse. Play proceeds in a clockwise fashion as each player moves a piece (or pieces) and relinquishes his turn. Players are allowed to “pass” if they do not want to play any pieces.

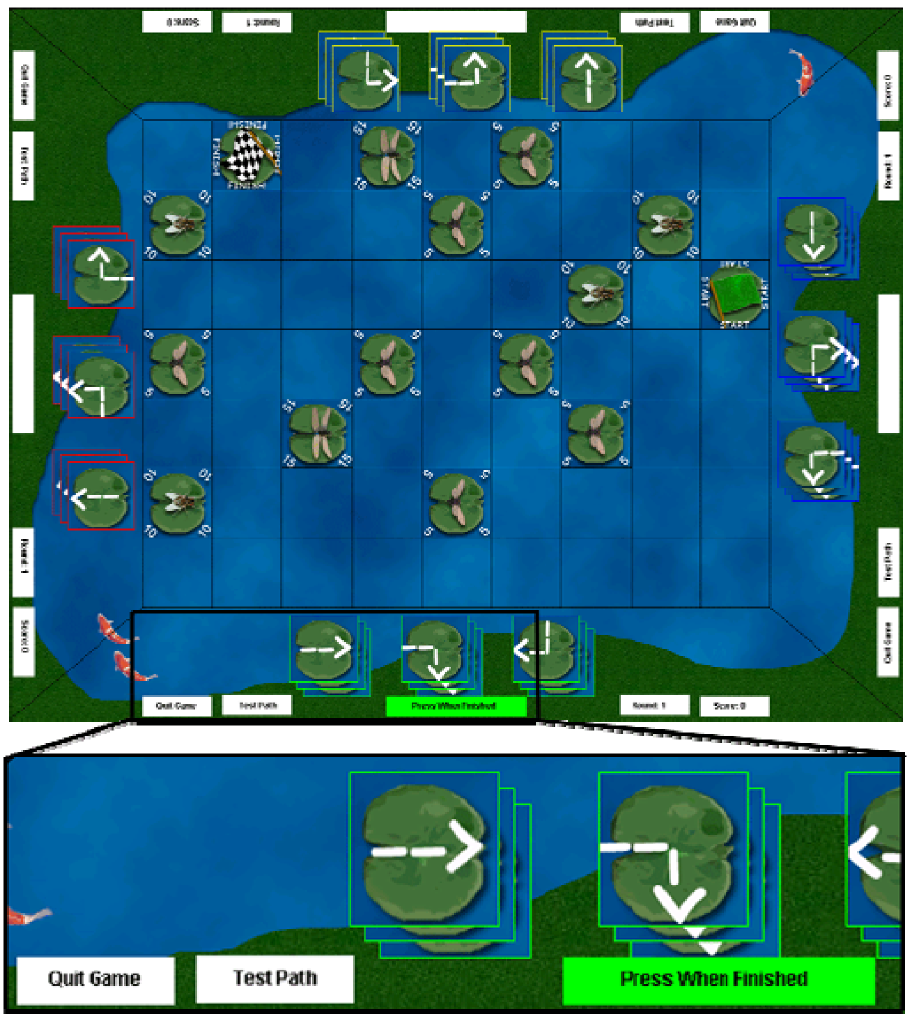


Figure 3.4: Screenshot of the improved design of the SIDES game. This version includes turn-taking and voting buttons for each player.

For Session 2, we decided to test the controlled access (players can only move their own pieces) and turn-taking features in combination, as this requires players to communicate more and to become coordinated in their attempts to create a solution. In our second evaluation, we examined how the adolescents from Session 1 and others from their class practice effective group work skills when playing a cooperative computer game under three conditions: (1) without enforced piece ownership and turn taking, (2) with piece ownership and turn taking enforced by a human facilitator, and (3) with computer-enforced piece ownership and turn taking.

3.7 Evaluation: Session 2

Session 2 focused on how these adolescents respond to computer- versus human-enforced rules and how elements of our design impact performance. The following questions guided this testing session:

- How do students respond to computer-enforced rules versus rules provided by a human facilitator?
- Do any aspects of the current design encourage or discourage effective group work with this audience?
- What is the role of a social skills therapist during a tabletop computer activity with this special needs population?

3.7.1 Method

To address these questions, we tested three variants of SIDES with two groups of four students, all from the same social therapy class. Group 1 consisted of four male students who participated in Session 1. The mean age of this group is 12.5 (stdev=0.58). The clinical diagnoses of individuals within Group 1 included two adolescent boys with AS, one with Apraxia, and one with Klinefelter's Syndrome. Group 2 consisted of four students, none of whom had played the digital version of SIDES yet, but three of whom had played the paper version in class. The mean age of Group 2 is 12.8 (stdev=1.5) and consisted of three male students and one female. In this group, two students were clinically diagnosed with AS, one with AS and ADHD, and one with High-Functioning Autism.

It is important to note that students in Group 1 had prior experience working with each other while playing the earlier version of SIDES during Session 1. In Session 1, these students experienced the "chaos" of playing without rules (i.e., no enforced piece ownership or turn taking). This experience gave them a benchmark to which they could compare their experience in Session 2. Group 2 had limited exposure to the game and minimal experience working with their set group of peers. For this reason, and due to the limited scope of our data set, we do not directly compare the two groups in Session 2. Instead, we treat the two groups as separate cases and seek to understand design implications based on the varying group dynamics and reactions to the activity.

The environment for this testing session, our university lab, was identical to that of Session 1. The two groups were presented with conditions as follows: Group 1: N, H, C, N and Group 2: N, C, H, N, where N = no rules, H = human-enforced rules, and C = computer-enforced rules. Each condition was presented as one round of play, where a round consisted of the group's successful construction of a complete path. In the N condition, students were presented with the basic version (similar to the version in Session 1, but with slight modifications to improve system performance) where no rules were enforced by the system and the therapist had limited involvement. The H condition again presented students with the basic version where rules were not enforced by the system, but under this condition, the therapist facilitated turn taking and enforced the "controlled access" of game pieces, only allowing students to move or play their own game pieces. In the C condition, turn taking and controlled access were enforced by the computer and the therapist had limited involvement in the activity, only providing occasional comments related to the group's strategy. Since Group 2 did not have prior experience with the computer version of SIDES, this group played the basic version without rules for approximately ten minutes to become familiar with the game and their teammates before beginning the conditions above. The same researchers who observed Session 1 also observed and took notes during Session 2. All game play and discussion was again videotaped for later analysis. Interactions with the interface were logged by the computer. After the testing session, students individually completed a questionnaire to compare the above conditions and then participated in a follow-up group interview.

We evaluate each group's performance individually and compare the reactions to the three conditions by each group in several ways. A critical part of our evaluation involves analysis of interaction over multiple rounds of play. The challenge these individuals face is not a lack of interaction so much as a lack of effectiveness in interactions (Bauer, 1996), therefore the effectiveness of verbal and non-verbal exchanges is an important indicator of cooperation by these adolescents. Our research team reviewed videos of both groups for Session 2 and independently coded player actions as follows:

- **Positive**: verbal agreement, agreement by making suggested play, encouragement
- **Aggressive**: verbal command, pushing, loud outburst or screaming, teasing
- **Non-Responsive**: ignore or dismiss idea without discussion, ignore or disregard therapist

We developed this coding scheme by consulting with psychiatrists and social skills therapists specializing in adolescents with AS, referencing the Diagnostic and Statistical Manual of Mental Disorders (DSM IV), and using our observations of play testing sessions to identify prominent themes. Inter-rater reliability between two researchers was above 85% (calculated with the Kappa statistic (Cohen, 1960)).

3.7.2 Group 1 Findings

According to our observations and student discourse, students in Group 1 exhibited an increase in positive language use as well as a decrease in the amount of aggressive behaviors over multiple rounds (Figure 3.5).

Based on conversational exchanges between group members, Group 1 as a whole performed best in the computer-enforced rules condition. Group 1 also demonstrated an improvement in conversation over the course of the trial and sustained this improvement in the final round without rules, the condition described as most difficult by students in Group 1. These students quickly adapted to the computer-enforced rules condition, becoming highly coordinated by skipping turns to get to a player who owned the piece necessary for the next move. Three out of four students in Group 1 rated the game as easiest to play when rules were enforced by the computer. Three out of four students in Group 1 also reported that they were most relaxed when rules were enforced by the computer. No students in Group 1 rated the computer-enforced rules condition as the most difficult version to play or as the condition they thought was most chaotic or most frustrating. Three out of four students in Group 1 said they worked together best during the computer-enforced rules condition and all four students reported that they worked together worst when there were no rules (condition N).

For this group, the computer-enforced rules condition encouraged cooperative group work because the structure helped prevent one dominant player from taking control of the game, as seen in the rounds without rules and in Session 1. With computer-enforced rules, each player had a clearly visible chance to "own" game play and make a contribution to the group's final product, one cohesive path. We suspect that for this group, the regimented group work experience was helpful in preparing students to work together in the final, unstructured round. During the debrief after Session 2, the therapist said to his students in Group 1, "You guys didn't even notice that in the last round you could touch each others' pieces and play in any order. You didn't reach across and take people's pieces like before. You kept working together."

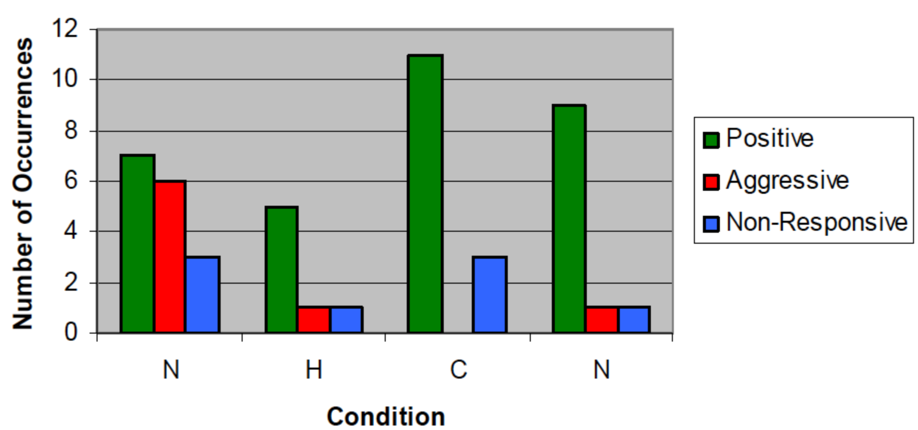


Figure 3.5: Analysis of interaction for Group 1. This group increased their positive behaviors in the computer-enforced rules condition and nearly sustained this rate of positive behaviors in the final no rules condition.

3.7.3 Group 2 Findings

In contrast to Group 1, all students in Group 2 stated that the game was easiest to play and that they worked together best when there were no rules. Three of the four students also indicated that they were most relaxed when there were no rules. The conversation analysis of Group 2 echoes the student questionnaire data. Group 2 exhibited more positive conversational exchanges and fewer aggressive behaviors in the no rules conditions (see Figure 3.6). Students in Group 2 sustained the same level of positive conversational exchanges and only slightly increased in aggressive behaviors over the four rounds. Group 2 indicated that the no rules condition was easiest and demonstrated conversation and behaviors that support their questionnaire responses. This group, however, did not indicate a majority opinion for the questions asking which version was most chaotic and most frustrating, but split their responses between the two conditions with rules. Responses to the condition under which the group worked together worst were also divided between the human- and computer-enforced rules conditions.

The challenge for students in Group 2 to work effectively in the structured conditions is partially due to the inflexibility of one player in this group. Brandon (age 11) consistently expressed skepticism about the team's solution and delayed the game by refusing to give up his turn even if he did not have any pieces to play. After observing Session 2, the therapist said, "I wish I could get the rest of my students to play this because it really gives me an idea of what's hard for each individual. Like with Brandon, I

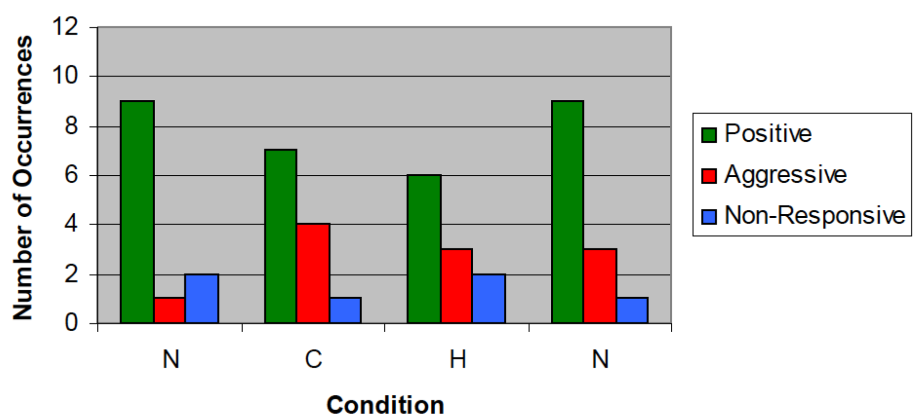


Figure 3.6: Analysis of interaction for Group 2. This group of students performed best in the no rules conditions.

had no idea he had such issues trusting other students until I saw him unwilling to give up his turn when the computer was enforcing turn taking.” For a group dynamic similar to that of Group 2, a “give” protocol in a turn-taking exercise can be problematic and detrimental to the group activity.

3.8 Discussion

SIDES provided an engaging experience for students who typically find group work extremely challenging and a source of anxiety. The students who played SIDES made a concerted effort to work with each other and remained engaged in the activity the entire time. Unlike traditional evaluations that do not target special needs populations, the learning that occurred during our testing sessions is impressive for these adolescents, as they quickly disengage from group activities when unmotivated by or uninterested in a task. Whether this behavior was a result of increased familiarity with the activity or their peer group, each group’s demonstrated interest in the activity and attempt to cooperatively devise one solution are successful outcomes for this user population and should not be underestimated. Even minimal strides in social skills development may not be visible until after months, even years, of exposure to therapy techniques. Our findings focus on initial exploration of tabletop technology as a tool for supporting effective group work with this audience and are not intended to capture sustained behavioral or long-term changes in these students, though this would be useful for future studies to address.

3.8.1 Design Lessons

We now present several design lessons from our development and evaluation of SIDES to help inform the design of future cooperative tabletop applications geared towards similar audiences.

Tabletop Technology as a Design Platform. We saw many benefits in the use of tabletop technology as the platform for an educational game for this audience. Interactive tabletop technology inherently supports social interaction and provides a shared experience for learners and educators, both of which are central to the learning process (Vygotsky, 1978). In an educational setting, it may be useful for the DiamondTouch table to distinguish between users and provide feedback on an individual basis. The specific hardware configuration of the table that allows user identification, however, also presents drawbacks for our user population and children in general. The DiamondTouch table requires users to remain capacitively coupled to the table (i.e., remain seated) to interact with the interface. This means that not only do students have to control their behavior and stay in their seat, but that serendipitous play by an observing student, therapist, or parent is not allowed. Furthermore, one of the shorter students had difficulty reaching across the table (which measures 107 cm diagonally) while remaining seated and therefore had trouble accessing any game pieces that were on the opposite side of the table. The ceiling-mounted projector that displays information on the interactive table surface also presents problems in that students must be extremely careful not to bump the table and misalign the projected image. The current state of tabletop technology, specifically the DiamondTouch, has promise as a medium for social computing applications for our target audience. It will be challenging, however, for tabletop computing experiences to impact larger special needs populations until more robust and affordable tabletop systems are developed.

Sensory and Motor Issues. The direct-touch input system of tabletop technology benefits our audience because it allows individualized and unmediated control over the interface, whereas the traditional mouse and keyboard setup has an additional barrier between the user and on-screen objects. All of the adolescents in our testing sessions manipulated the virtual game pieces well and demonstrated effective control over the interface. Our evaluation did not reveal any sensory or motor issues related to our user group that inhibited the functional use of tabletop technology. All participants were high-functioning and none had motor coordination difficulties that would significantly impact the use of a traditional computer workstation with a keyboard and mouse. Adolescents with Autism Spectrum Disorders vary in sensory tolerance and motor ability levels, so it is important to evaluate an adolescent's ability to use tabletop technology on an individual basis.

Human- Versus Computer-Enforced Rules. Computers provide reliability and consistency in rule enforcement which is particularly useful for these adolescents who find comfort in predictable rules and environmental conditions. Rule enforcement by a human moderator in SIDES and traditional board games can often be subjective and add challenge to an already difficult task, making computer-enforced rules a compelling aspect of tabletop games for this audience. The students' therapist commented, "These kids generally do better with rote, impersonal, nonsocial instructions. That's why they do well with computer games. There's no variance, so they don't have to worry about social conventions or social rules." When asked to compare how his students performed in the conditions with computer-enforced rules and human-enforced rules, the therapist replied:

> It's hard because I thought that they did better without me and my input. I tried to get them to think about strategy, but there was so much stimulus and enjoyment in the game that they didn't listen to me! They had to respond to an adult when I was facilitating it. The computer rules version eliminates one social interaction that they otherwise would have to attend to... Just listening to the game, which is more objective, made playing easier.

Embedded Structure. A mechanism to decentralize control, voting buttons in the case of our design, can encourage users to collectively own the process of finalizing and testing a solution. While one player often suggested that the group test the solution, the voting buttons required all group members to come to consensus, encouraging discussion and compromise around changing the state of the game. Our findings also indicate that a turn-taking mechanism (tested with a "give" protocol in Session 2) can reduce an individual's ability to dominate the tabletop activity and take control from others (as illustrated by Group 1 during Sessions 1 and 2). Piece ownership (players could only touch and move their own pieces) provided a way for each student to play a critical role in contributing to the group's solution and creating the path. For example, Brad, a more reserved student in Group 1, received several critical game pieces during various rounds of play. Piece ownership in this case provided him with a chance to contribute key elements to the final product and become more involved in the activity.

Computer-enforced turn-taking combined with restricted access to game pieces necessitates that groups achieve a coordinated state where each player contributes their resources in an orderly fashion to create the final solution. This level of structure worked well for Group 1 but became problematic for Group 2 who was held up by one player who was unwilling to negotiate and trust his teammates' strategy. This student struggled with controlling his frustration when the computer restricted piece movement and he consequently disrupted game play for others by refusing to give up his turn. In this case, the embedded game structure provided additional challenge and hindered the second group's performance. At that point, it would have been interesting for the therapist to try rule variations, gradually reducing structure to determine the most appropriate level of support for this group. In future designs and play sessions, we envision that the therapist is able to adjust the type of rules and how rules are enforced so that students experience a gradual increase in difficulty that is customized to their learning needs.

Need for an Adult Moderator. Involvement of the students' therapist is critical to delivering a customized, cooperative group work experience for these adolescents. Digital tabletop activities, SIDES in particular, are not yet intended to stand on their own as a tool to support effective group work with this audience. The social skills therapist plays a central role in facilitating tabletop activities; his or her presence is required during the session to control behavior and attend to higher-level group work issues. After each play session, the therapist also plays an important role in grounding the learning experience in the social skills concepts discussed in class, which can seem extremely abstract to these students. Through discussion of the activity immediately following game play and even up to weeks later in class, the therapist helps students reflect on the activity and tie abstract classroom topics into this shared real world experience. "The key is to give [the students] experiences to trust themselves, trust their abilities to interact so that generalizes to interacting with other kids in other settings... The goal is generalizing the experience," explained the therapist. SIDES provided a rich and meaningful shared experience that the social skills therapist was able to leverage during classroom discussions, thus revealing the potential for supplementing current classroom activities with an exciting and supportive tabletop computer gaming experience.

Challenges in Participatory Design. Participatory design was critical to designing a motivating experience for this audience and exploring the potential for tabletop computer applications to facilitate effective group work among these adolescents. However, involving adolescents with AS, their parents, and therapists in the design process presented challenges. We share some of our experiences in light of other research on participatory design with special needs groups to inform future work within this discipline.

As previous research on participatory design with special needs groups indicates (e.g., Moffatt et al., 2004), getting cooperation and assistance from existing groups who cater to the target population and gaining practical experience with the target population was invaluable to our design process and system evaluation. Identifying and receiving entry into an existing group of adolescents with AS, specifically a middle school classroom, was difficult. Obtaining permission to observe this classroom and involve students in our research was an extremely delicate subject and required multiple layers of approval. Many of the students we observed and interviewed had not been informed by their parents of their diagnoses. We faced issues receiving parental consent and initial "buy in" on our project for this reason. It took months to build rapport with this group of students and their parents. This required talking extensively with parents before student interviews and allowing parents to review our interview protocol (per their request) to ensure that nothing we covered would upset their child.

Once we received school, teacher, parental, and student permission, we had another set of challenges to overcome with conducting observations and interviews with this small group of students; we would often arrive at the classroom to find that the student with whom we had worked for several weeks to get an interview scheduled and approved would be in a "bad mood" according to teachers and unwilling to cooperate. Traditional one-on-one interview situations were less successful than we had hoped due to communication and behavioral issues with students. We learned to be flexible in our methods without compromising data validity. For example, we shifted from conducting single person interviews to a group format where students could discuss ideas.

Wu et. al. (2005) recommend implementing specific techniques to directly support the challenges inherent to the target user group. For our user population, feeling comfortable in a new environment and avoiding over-stimulation were primary challenges. To address these issues, we spent months building rapport with these students in their classroom environment, encouraged parents to attend the testing sessions at our university, and structured our testing sessions to encourage students to only focus on one task – cooperatively solving the computer puzzle in front of them. Parents were allowed and encouraged to attend testing sessions, as this helped them see the value in our research first hand as well as calm their child if the student became anxious in the new environment.

These day-to-day challenges of conducting participatory design with this audience helped our research team understand how the design of a motivating cooperative skill-building activity could have a far-reaching impact for this population, including students, parents, and social skills therapists.

3.8.2 Overall Impact

Our work with SIDES reveals the potential for supportive social entertainment applications implemented with tabletop technology to address group work issues among special needs populations. The affinity for technology individuals with AS describe, combined with the ability of computer technology to enforce basic game rules (thus freeing up therapists' time to deal with higher-level group issues), and the flexibility of computer games for adapting content and difficulty level, make tabletop technologies more compelling for this user population than traditional board games. Regarding the students' experience, the therapist commented,

> It's something they enjoyed doing, so it's not like a lesson where you're teaching them something in traditional lesson form. With the game they're just learning these skills by doing something fun. It's like you're sneaking in learning without them knowing it... It's great that they can feel confident and comfortable while working with each other because it's not torturous. These students didn't even see the activity as learning to work in a group.

Helping students build confidence in their social abilities is another benefit of playing SIDES. For Brad, participating in the testing sessions was an experience far beyond just learning to work in a group. "[Brad] is a kid who has been tormented and terrorized by other kids in his class. For him to be able to participate and feel like he's part of the group and accepted was great. He probably enjoyed it more than anyone because his existence was validated through the shared activity," commented the therapist. On both an individual and class-wide level, we observed the positive effects of situating a topic that is traditionally difficult for this group of students, learning effective group work skills, in an exciting and comfortable context: playing a cooperative tabletop computer game.

3.9 Conclusion

We have presented a design case study of a cooperative tabletop computer game for a special needs population. SIDES provides adolescents with Asperger's Syndrome with a positive experience through which they can develop effective group work skills and build confidence in social interaction. This work provides a starting point for thinking more broadly about user populations and computing scenarios that can benefit from the social computing experience provided by tabletop technology. Through our design process, we thought critically about how the unique social benefits of tabletop technology could benefit this population and crafted an application to support the needs of this group. We believe cooperative computer games are a new paradigm for teaching effective group work skills in a meaningful way and that tabletop technology is a promising tool for facilitating cooperative gaming experiences for this special needs population, as well as the general public.

Chapter 3, in part, is a reprint of the material as it appears in the Proceedings of the 2006 ACM Conference on Computer-Supported Cooperative Work (CSCW). Piper, A.M., O'Brien, E., Morris, M.R., and Winograd, T. SIDES: A Cooperative Tabletop Computer Game for Social Skills Development. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Shared Speech Interface

This chapter describes the design and evaluation of Shared Speech Interface (SSI), an application for an interactive multitouch tabletop display that supports medical conversations between a Deaf patient and a hearing, non-signing physician. We report on an evaluation that compares conversation when facilitated by: (1) SSI, (2) a human sign language interpreter, and (3) both SSI and an interpreter. Results from this study indicate that SSI is a more private and independent alternative for communication than a human sign language interpreter.

4.1 Introduction

This chapter presents the design and evaluation of Shared Speech Interface (SSI), an application for an interactive tabletop display that augments communication between a Deaf[1] patient and a hearing, non-signing medical doctor. Currently, medical facilities provide a sign language interpreter to facilitate communication between a Deaf patient and hearing doctor. Deaf patients may need to plan ahead to ensure that an interpreter is available. For most patients, privacy is a central concern. Depending on their comfort level with interpreters and the topic of discussion, some Deaf patients prefer to use an alternate communication channel, such as email or Instant Messenger, to discuss medi-

[1]The word "deaf" with a lowercase "d" refers to the physical condition of hearing loss. The word "Deaf" with an uppercase "D" is a cultural label for the community of people who identify ASL as their native language. All of the participants in this study identify themselves as Deaf so I adopt the same convention here.

cal issues with their physician. Increasing privacy and independence in communication is one design goal for the system described in this chapter.

Tabletop displays with speech recognition have the potential to facilitate medical conversations between Deaf and hearing individuals. Consultations with physicians often involve discussion of visuals such as medical records, charts, and scan images. Interactive tabletop displays are effective at presenting visual information to multiple people at once without necessarily designating one person as the owner of the visual. Taking notes while meeting with a physician is problematic for Deaf individuals because it requires attending to the doctor's facial expressions, the interpreter's visual representation of speech, and notes on paper. SSI allows the doctor and patient to engage in face-to-face interaction while viewing an interactive transcript of conversation on the shared interactive surface. The system incorporates keyboard input by the patient and speech input by the doctor, allowing the physician to speak and gesture as they discuss medical details and visuals with the patient. Importantly, speech recognition allows the Deaf patient to read the doctor's lips and use SSI as an augmentative communication aid.

Our work on SSI provides practical experience designing a shared communication device for users with different hearing abilities. We examine the challenges of representing speech visually to multiple users around a tabletop display. Finally, we provide design guidelines for using dictated speech with shared display systems and discuss the implications of our work on multimodal tabletop displays for the broader CSCW community.

4.2 Background

Loss of hearing is a common problem that can result from noise, aging, disease, and heredity. Approximately 28 million Americans have significant hearing loss, and of that group, almost six million are profoundly deaf (NIDCD, 2008). A primary form of communication within the United States Deaf community is American Sign Language. ASL is not a visual form of English; it is a different language with its own unique grammatical and syntactical structure. Sources estimate that ASL is the fourth most commonly used language in the United States (NIDCD, 2008). While ASL is widely used in the United States, no one form of sign language is universal. Different countries and regions use different sign languages. For example, British Sign Language is different from American Sign Language, although both countries have English as their official and primary spoken language.

There is great variability among individual needs and abilities within the Deaf community. Not all individuals who are born Deaf are raised in the aural tradition where they are taught to speak, read, and write in English or another spoken language. Others who suffer from late-onset hearing loss typically have fully developed vocal abilities but are not fluent in ASL and do not know how to read lips. This range of individual abilities and needs has lead to a gamut of techniques for communicating with hearing people.

4.2.1 Adaptive Techniques

For the Deaf population proficient in a spoken language such as English, writing has long been a central form of communication with hearing people. Deaf individuals may read lips, use handwritten notes, and gesture as part of communication. Telephone use was impossible for the Deaf community until the invention of the Teletype and Text Telephone (TTY), a typing-based system that transmits individual lines of keyboard entry over phone lines. Adoption of the TTY and eventually the personal computer made typing in a written language such as English an essential mode of communication for the Deaf community. In recent years, the invention of webcams and increasing Internet bandwidth gave rise to communication through video chat with other ASL speakers and ASL interpreters.

ASL interpreters play a central role in enabling face-to-face communication between Deaf and hearing individuals. For the Deaf population fluent in ASL, communicating through an interpreter is an optimal choice for many situations. Interpreters, however, are expensive and not always available. Furthermore, though interpreters are bound by a confidentiality agreement, the presence of a third person in a highly private conversation may reduce a Deaf person's comfort and inhibit their willingness to speak candidly.

4.2.2 Technologies for Deaf People

As early as 1975, researchers began investigating how cooperative computing environments, such as early forms of Instant Messenger, could facilitate communication between Deaf and hearing individuals (Turoff, 1975). More recently, HCI researchers have examined how mobile devices, tablet computers, and video conferencing technologies can augment communication for Deaf individuals. Schull (2006) presents communication via a browser-based client on multiple collocated laptops that allows a Deaf and hearing user to access a common browser window and share real-time chat information. iCommunicator (2008) is a commercial product that enables communication in a similar way. Few systems, however, facilitate shared face-to-face communication experiences. The majority of work focuses on single-user interfaces for distributed applications. For example, MobileASL enables two signing individuals to communicate with each other over cell phones with real time video (Cavender et al., 2006). Scribe4Me, a mobile sound translation tool, enables Deaf individuals to request a transcription of the past 30 seconds of audio in their environment (Matthews et al., 2006). Matthews et al. (2005) use peripheral displays to visualize various channels of auditory information for Deaf individuals. The Facetop Tablet project examines visual attention problems that Deaf individuals experience when trying to view a presentation, watch an ASL interpreter, and take notes (Miller et al., 2007). This project presents a viable approach for helping Deaf individuals follow a conversation with hearing individuals but it does not explicitly help Deaf individuals communicate or become active participants in the conversation. Related research also examines enhancing communication for Deaf individuals through gesture recognition with computer vision and wearable computers (e.g., Starner et al., 1997).

While these solutions address various communication challenges for Deaf individuals, none provide a shared communication system that addresses speaking and listening needs between Deaf and hearing people. Furthermore, many of these communication technologies only provide an interface and feedback to one user. In contrast, SSI leverages the cooperative nature of interactive tabletop displays to enable a shared, co-constructed communication experience between users with varying hearing and speaking abilities.

4.2.3 Tabletop Displays

There is a growing history of research on tabletop displays and their utility for supporting group work. This body of research examines cooperative group work around tabletop displays with hearing populations and is the foundation for our research. We leverage techniques from research on social protocols around digital tables (Morris et al., 2004b), the notion of shared and private spaces on tabletop displays (Scott et al., 2004), and how to present textual, pictorial, and auditory information to users (Morris et al., 2004a; Ringel et al., 2004b).

Research on multimodal tabletop interaction, specifically integrating speech with touch input, influences this work. Tse et al. (2008, 2007b, 2006) investigate speech commands on a tabletop in group design and gaming experiences. The Reflect project (Bachour et al., 2008) uses a traditional table embedded with LEDs to display attributes of conversation, such as the percentage of time each person spends talking. SSI builds on this research by examining a novel application and implementing an interactive conversation transcript generated by dictated speech.

4.3 Design Process

We employed a human-centered design process, involving members of the Deaf community and domain experts in all aspects of the design and evaluation of SSI. This included regular interviews and prototype reviews with Deaf individuals, linguists studying Deaf culture and sign language, and medical professionals who communicate regularly with Deaf individuals.

4.3.1 Understanding Communication Challenges

As previously mentioned, sign language interpreters are an important link enabling communication between Deaf and hearing individuals. The physical presence of an interpreter, however, may inhibit a Deaf patient from candidly discussing private medical issues. Other challenges in interpreter communication include accuracy of interpretation and challenges for the Deaf individual when attending to multiple channels of visual information.

Karen[2] is a Professor of Communication who studies Deaf culture and sign language. She was born Deaf and both of her parents are Deaf; however, she was raised in the aural tradition by attending hearing schools. Karen is exceptionally skilled at reading lips and speaking. She does not need an interpreter for most situations. In an interview she mentions her Deaf husband's privacy concerns, "My husband for example, lip reads his doctor... he doesn't want an interpreter. If he needs an interpreter, he wants me to come." Karen describes another situation and reiterates the need for a more private and independent communication alternative:

> I know one situation where a therapist needed to see a Deaf person on an emergency situation and [the patient] wasn't comfortable going through an interpreter. So they had two TTYs. They actually took the TTYs and were typing back and fourth. But the problem is it only had one line at a time. It was hard to remember what the issue was if you kept losing it.

When an interpreter is provided by a medical facility, the Deaf individual usually does not have a choice about who will interpret. The Deaf community is well-connected in our city, and one Deaf woman said that she would not feel comfortable going through an interpreter that she knew well or going through a male interpreter. She emphasized the need for a better option when a Deaf patient is not comfortable with the interpreter.

Beyond issues of privacy and autonomy, ensuring accuracy of communication with an interpreter is critical. Although medical interpreters undergo special training, it is often the case that interpreters are not experts on the content they are interpreting. For example, one Deaf individual described having an interpreter in Biology class who knew little about Biology. This made understanding the interpretation extremely difficult and impacted her ability to learn in the classroom. When it comes to health care issues, receiving accurate and full interpretation is essential. Dr. Stevens is a clinical professor at our university hospital. She runs a program to teach medical students about Deaf culture and how to create a Deaf friendly medical environment. Dr. Stevens comments on the challenges of relaying information through an interpreter:

> We're assuming that interpreters have a lot of medical information, and they may not. They may be miscommunicating. I always think about the doctor who told the patient, take three of these a day, and the interpreter didn't know to explain take one morning, lunch, and dinner.

[2] All names of research participants have been changed to preserve anonymity.

Facilitating conversation with an interpreter also creates challenges for Deaf individuals to share their attention between the conversation and note-taking. Dr. Stevens explains:

> With an interpreter... the patient has no ability to make a record of what's being said because the eyes are on the interpreter and they can't be on your paper, on the interpreter, on the doctor getting your emotions. What's your body language saying? You're showing me my knee but I'm looking at your face to see how bad this really is... So most of us go home with notes written down, but if you're Deaf, it's hard to do both.

The goal behind our design for SSI is to explore an alternative to human interpreters as well as to augment conversation that occurs through an interpreter.

4.3.2 Technical Implementation

We prototyped SSI on a MERL DiamondTouch table (Dietz and Leigh, 2001) using the DiamondSpin toolkit (Shen et al., 2004). The DiamondTouch table is a multiuser, multitouch, top-projected tabletop display. Users sit on conductive pads that enable the system to uniquely identify each user and where each user is touching the surface. Our system enables conversational input through standard keyboard entry and a headset microphone. The audio captured from the microphone is fed into a speech recognition engine (currently this is Microsoft Windows' default recognizer, but our application is easily adapted to any recognizer). SSI uses the Java Speech API (2008) and CloudGarden API (2008) to interface with the speech recognition engine and send converted speech-to-text into the main application running on the DiamondTouch table.

Because understanding and analyzing conversation becomes increasingly complex as the number of speakers increases, we implemented the first version of SSI for two users only, one hearing user and one hearing-impaired user. We first investigate whether tabletop displays are an effective communication tool for this basic case of two users prior to expanding to larger group work scenarios. The discussion section at the end of this chapter provides ideas for expanding and adapting the SSI technology for more than two users.

4.3.3 Interface Design

We use a tabletop display because it allows the doctor and patient to maintain a face-to-face working style (Rogers and Lindley, 2004). For both participants, eye contact and facial expressions are important communication channels. Certain positions afford easier eye contact and information sharing, especially when speakers are positioned around a table. Sitting across from someone gives direct access to facial expressions but makes sharing textual information problematic, although some solutions for reorienting and rotating text have been explored (Ringel et al., 2004b). While sitting side-by-side makes sharing textual information much easier, the critical activities of making eye contact and reading lips becomes challenging. With SSI, users sit next to each other but at the corners of the table so that they can make eye-contact, read lips, and share text-based information.

As each person contributes to the conversation, either by speaking into a microphone or typing on the keyboard, their speech appears on the tabletop display in front of them. We refer to these fragments of conversation as "speech bubbles." Speech bubbles are color-coded by user and moveable around the display. In ASL, speech has a visual and spatial element, and signers often point back to where a previous gesture occurred. The persistent nature of speech bubbles in our design may be an analog for the spatial nature of ASL signs. Seating users at angled positions makes uniformly orienting speech bubbles towards the bottom of the display a natural choice. In the first version (see Figure 4.1) we present speech bubbles close to the user who entered the speech but not in an ordered fashion. We want to examine how people organize speech bubbles on the display and determine whether preserving turn-taking is important.

With our design, both users could have easily entered conversation through keyboards. We hypothesized that it was important for the physician to speak naturally to the patient. The patient could then attend to the physician's body language and read their lips. The doctor's facial expressions and body language are masked when speech is entered through a keyboard. As noted by Dr. Stevens, reading the physician's body language is an essential part of communication in medical conversations.

While speech recognition engines are constantly improving, transcribing natural language into text is still problematic. It is critical that the speaking user can control the speech that the system displays. To enable this, SSI provides the speaking user with three "best guesses" of their speech from the recognition engine (see Figure 4.1). The user touches the phrase that matches their intended speech and the phrase appears on the interface as a speech bubble. While this requires an extra step for the speaking user, this model provides greater accuracy in speech recognition and prevents unintended conversation from appearing on the display. The system also enables conversational repair. Users delete previous parts of conversation by dragging speech bubbles to the trashcans in the top corners of the display. It is important that users can easily correct miscommunications and remove unwanted or incorrect speech.

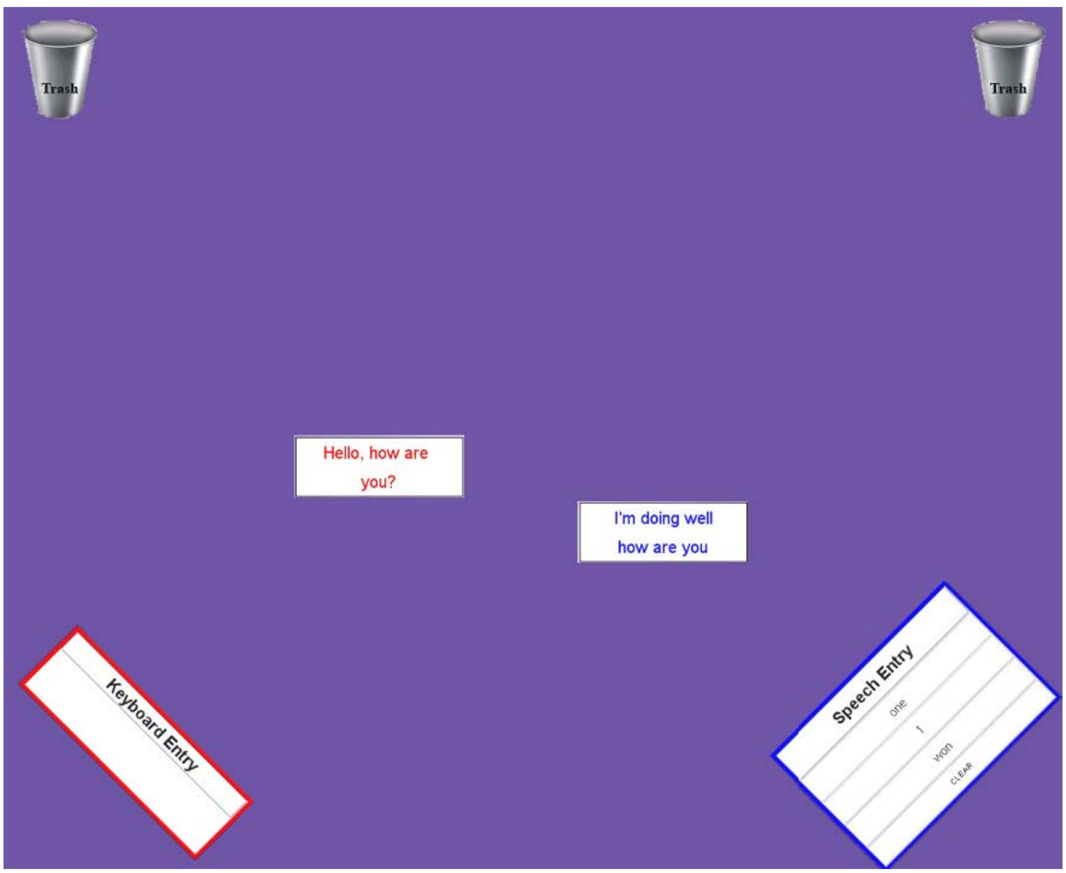


Figure 4.1: First interface design. Users sit next to each other but at the corners of the display. Speech bubbles are color coded by user. The doctor (seated on the right) has a dialogue box that allows her to select one of three phrases that best matches her speech.

4.3.4 Prototype Review

As part of our design process, eight people reviewed this early interface design (as shown in 4.1), four of whom are Deaf and the other four are actively involved in the Deaf community. An interpreter was present to facilitate communication between the Deaf participants and hearing participants and researchers. Our questions at this stage of design addressed user position at the table, text size and color, speech bubble behavior, and the use of keyboard entry. This preliminary feedback revealed several key issues about the design of SSI. We discuss these and explain how they influence subsequent designs and evaluation.

Overall User Interface Design. We found that positioning users at an angled position around the table works well. One Deaf person said, "It feels too much like teacher and student when people sit across from each other." Maintaining eye contact and reading shared speech bubbles is not a problem with the current configuration. Reviewers said that the text size and colors work well, but because the speech bubble text is fairly large, the display clutters quickly. Displaying medical images in the background is a positive aspect of this design, especially for stepping through the slides of an MRI or indicating fracture points in an x-ray image (see Figure 4.2).

Organizing Conversation, Managing Clutter. In the first interface design, the display fills up quickly with speech bubbles, even with a trashcan to delete unwanted conversation. Reviewers said the trashcan is useful for correcting misunderstandings and removing unwanted speech bubbles (especially large speech bubbles). However, there is still a need to organize conversation and several people suggest being able to create a new page.

We considered zooming and scrolling interfaces to increase screen real estate, but instead we introduced a simple tabbed design that leverages users' knowledge of modern web browsers (see Figure 4.2). One Deaf man said, "The tabs are a great way to know what you're going to work on and how you're going to move forward. You can go back and refer to something without having to search for it...the tabs are a nice way to do that." In response to his comment, a Deaf woman said, "exactly, I think it's good because it's not overwhelming...it's very Deaf friendly. It's very visual." Overall, reviewers like the tabbed design and find it easy to understand.

In an early design, we also presented speech bubbles at an arbitrary location in front of the user who contributed the speech. Reviewers said that it is difficult to anticipate where speech would appear and that it is important to preserve turn-taking. The design now preserves turn-taking by presenting speech bubbles in a linear pattern, offset and color coded to indicate the speaker (see Figure 4.2).

Pre- and Post-Discussion Use. According to Dr. Stevens, doctors often know several key discussion points before they enter a meeting with a patient. She said that it would be extremely helpful to have the nurse enter talking points to structure and pace the conversation. These could be available on the doctor's side of the interface. Dr. Stevens also mentioned that it would be good to add labels to the tabs to show the topics for discussion (e.g., "welcome" and "update since last visit"). Reviewers unanimously want the dialogue to persist after the appointment. They mentioned saving the conversation for the next visit, printing it to take home and share with family, and receiving a copy of it via email.

Diversity within the Deaf Community. One important realization from the prototype review is that there is great diversity within the Deaf community and that our system would work better for certain subpopulations. Members of the Deaf community thought that SSI would work well for Deaf individuals who feel comfortable using English and for individuals who are Hard of Hearing. They said SSI would be problematic for Deaf individuals with low English literacy or low confidence in English communication.

4.3.5 Design Modifications

Based on feedback from the preliminary evaluation, we kept the text size and font the same, added a feature to display speech in an ordered fashion that preserves turn-taking, and proceeded with the tabbed interface design. We made a slight modification to the look of the speech bubbles by adding a tail on each bubble that indicates its author. After testing speech recognition with multiple people, including medical professionals, we added a button to bring up a virtual keyboard so the doctor can type the words that are difficult for the speech system to recognize.

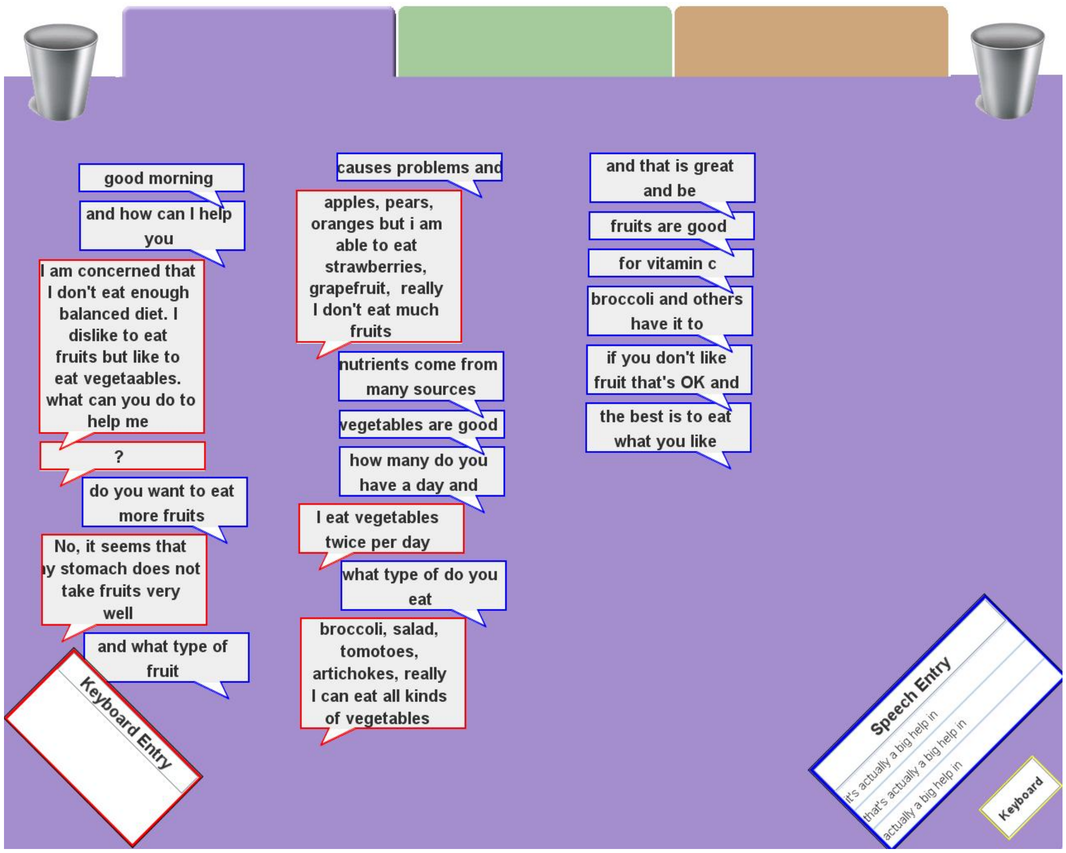


Figure 4.2: Revised interface design that preserves turn-taking in conversation by initially ordering speech bubbles. Tabs across top of display provide additional conversation space. Images to support conversation can be preloaded into the system under one of the tabs.

4.4 Evaluation Method

We evaluated SSI through a laboratory experiment with eight Deaf participants (mean age=33, stdev=11.4, range=[22,52]; 3 males) and one medical doctor (age=28, female). All eight Deaf participants were born Deaf or became Deaf before the age of one. Three participants identify English as their native language and five identify ASL. All participants are fluent in ASL and proficient at reading and writing in English. In this experiment, each Deaf participant discussed sample medical issues with the same doctor. This resembles the real-world scenario where one doctor has similar conversations with multiple patients throughout the day. None of the participants had experience with a tabletop display prior to participating in this evaluation. The evaluation took place in our university laboratory around a DiamondTouch table. Each session was video taped by two cameras from different angles to capture participants' interactions with each other and the digital table. Also, the computer unobtrusively recorded all user interactions with the tabletop surface. Three researchers observed the testing sessions and took detailed notes.

Procedure. The first step was for the doctor to train the speech recognition engine to recognize her voice. This took approximately 20 minutes. At the beginning of each testing session, the Deaf participant and doctor performed a brief period of training together to get to know each other and adjust to the task. Then the patient and doctor discussed a medical issue using SSI (Digital Table), a human interpreter (Interpreter), and both SSI and the interpreter (Mixed). At the beginning of each conversation, the Deaf participant received a discussion prompt in English text about a medical topic (e.g., nutrition). Each discussion prompt had a corresponding medical visual preloaded into the system under the third tab. The other two tabs were blank conversation space. A paper version of the visual was provided for the Interpreter condition. Medical professionals worked with us to ensure that the discussion prompts reflected authentic conversations that might occur in a normal patient interaction but whose content did not require participants to discuss information that might be too personal. Both the order of conditions and discussion prompts were randomized between subjects. Each conversation lasted between seven and nine minutes. Finally, the Deaf participant completed a questionnaire about their experience.

4.5 Findings

Overall, digital tables are a promising medium for facilitating medical conversations involving Deaf individuals. The most striking contrast is between the Digital Table and Interpreter conditions, and we focus discussion on those two conditions. We observed an interesting use of gesture to augment communication in the Digital Table condition and several differences in conversation between the conditions. Questionnaire data indicates that our application is good for private conversations and enables independence. However, the interaction overall is limited by the state of the technology and certain aspects of communication are challenging in practice. Specifically, imperfections in the speech recognition engine made conversation in the Digital Table condition considerably slower than in the Interpreter condition.

4.5.1 Conversation and Gesture

There were several key differences in communication between the Digital Table and Interpreter conditions. The Digital Table condition allowed for asynchrony in communication, whereas the interpreter acted as a broker of conversation and thus encouraged synchronous interactions. Dialogue in the Interpreter condition was verbose and elaborated, while speech in the Digital Table condition was more concise and telegraphic in nature. We observed balanced participation in the two conditions, and the doctor and patient each contributed to about half of the conversation.

Slower communication in the Digital Table condition was due to latency in the speech recognition process. The system took one second to determine and display the speech in textual form. Then the interface required the doctor to tap on the phrase she wanted to add. The best speech recognition result occurred when the doctor broke her natural speech into short phrases, but this also slowed communication. On some occasions the doctor made two or three attempts before the system accurately recognized her speech. While speech recognition was problematic, it provided ancillary benefits such as allowing the doctor to gesture while speaking and the patient to read the doctor's lips and facial expressions.

In the Digital Table condition, we observed that gesture by both the doctor and patient played an important role in augmenting and ensuring successful communication. Importantly, the collocated, face-to-face nature of the digital table allowed participants to provide feedback to their partner about their state of understanding through deixis, gaze sharing, and head nodding. Participants strategically moved speech bubbles in front of the other user to get their attention and pointed between bubbles to make a connection to previous speech (Clark, 2003). We also noticed a pattern in which one participant moved or pointed to an object on the interface and then one or both participants nodded to confirm their understanding. This pattern occurred frequently in the Digital Table condition. We elaborate upon these observations through a detailed analysis in Chapter 5.

4.5.2 Affordances of Digital Space

With SSI, the digital table transforms the ephemeral nature of speech into a tangible and persistent form, creating affordances that are not available in traditional conversation. We observed interesting behaviors with the speech bubbles because of their form. When a phrase was added to the display that referred to a previous utterance, the "owner" of the speech bubble often moved the new phrase close to the previous utterance. In conversation, the speaker must help listeners understand a reference to a previous utterance through context and explicit referencing. The digital table allowed users to reference previous conversation by placing new speech near an existing speech bubble. The persistent nature of speech with the digital table allowed participants to review their conversation. We observed both the doctor and patients looking back over their previous conversation. The doctor said "it was good to look back at what I had covered with that particular patient," and explained that "it would be helpful because it is not uncommon in medicine to have very similar conversations with different patients throughout the day."

4.5.3 Presence of an Interpreter

We noticed differences in how patients attended to the doctor when the interpreter was present. Deaf participants made some eye contact with the doctor when they signed but then shifted their gaze to the interpreter when the doctor began speaking (Chapter 5 details this early observation). In the Digital Table condition, participants typically looked at the doctor when she was speaking and then looked down at the display. In the Mixed condition, we observed a pattern in which the patient watched the interpreter sign and then looked down at the display to read the English version. Several participants explained that seeing the doctor's speech in both ASL and English was helpful. The interpreter also found benefit in having the digital table present:

> What was nice for me as the interpreter was to have the printed word. When I didn't know how to spell something, especially in medical situations, if the printed word is there... I can point to it... so that about the communication board is very attractive.

4.5.4 Communication Preferences

At the end of the laboratory study, each Deaf participant completed a questionnaire about their preferred method of communication. Figure 4.3 summarizes participant responses.

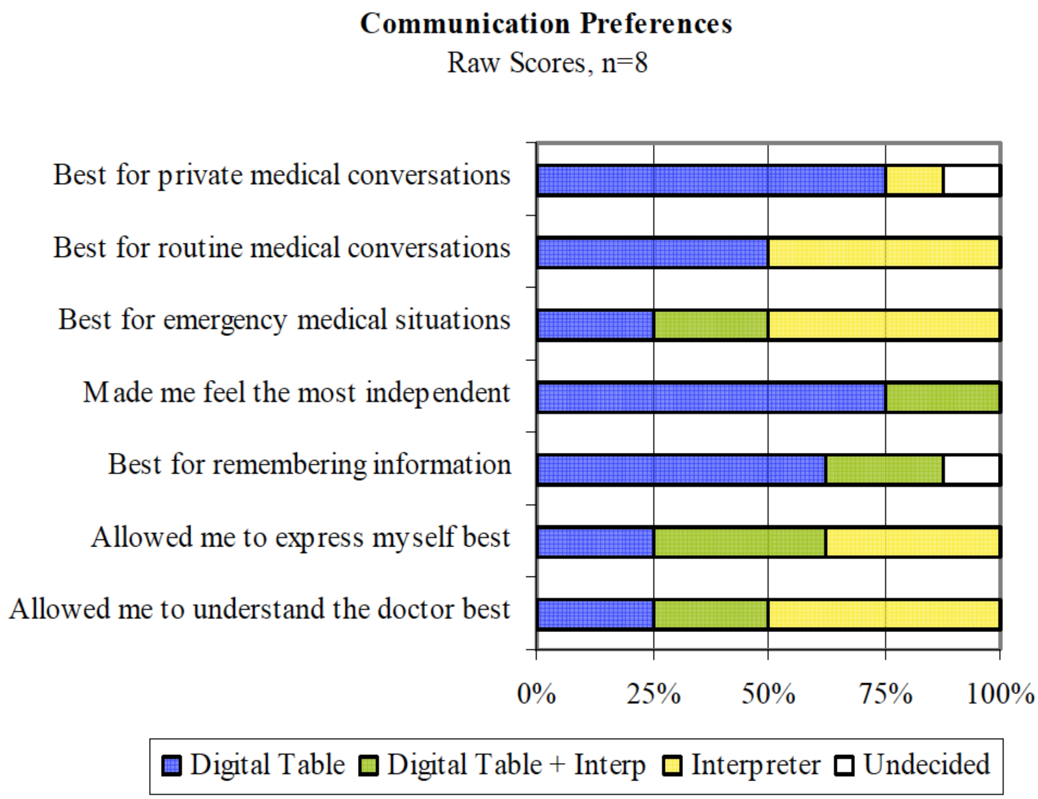


Figure 4.3: Questionnaire data from Deaf study participants regarding preferences for communication.

Privacy. Six of eight Deaf participants reported that the digital table alone was best for private medical conversations. Cathy said, “the digital table is best for very private conversations, but using an interpreter in a private conversation depends on whether or not I know the interpreter.” Sharon explained, “for other meetings, like a work situation or job interview, I would prefer to have an interpreter. But for personal meetings, like with a lawyer, doctor, or specialist, I prefer the digital table.” Jesse also said the digital table is good “if a client feels they can’t confide with an interpreter present.”

Independence. Six of eight participants reported that the digital table alone made them feel the most independent. Amber explained, “I don’t have to wait for an interpreter. It saves time.” Cathy, on the other hand, said, “it’s true, I did not need to rely on an interpreter, but sometimes independence isn’t exactly what I want. I value smoother conversation.” There is a tradeoff involved with using the digital table: conversation may be slower, but the patient has autonomy and privacy.

Remembering information. Although we do not have data to judge this, seven participants indicated that using the digital table would be helpful for remembering information. Participants said “it provides a record that I could go back and look at” and “it’s all documented.”

Understanding the doctor. Half of participants stated that including the digital table helped them understand the doctor best. Jesse explained, “the table could be used to clarify words that the interpreter may not understand or comprehend.” Similarly, Mark said, “the table showed the exact words the doctor used.” Alex said her preference “depends whether the interpreter is fluent and sharp. The table is better if the interpreter is bad. In that case I would prefer to type for myself.”

Speed of conversation. In follow-up discussion, several participants said they preferred the interpreter when speed of conversation was critical. Sharon said “the table will work only when both parties are patient.” Amber explained, “in an emergency I won’t have time or energy to type.” There are cost-benefit tradeoffs between communicating through a quicker channel (the interpreter) versus a private and independent channel (the digital table).

4.6 Discussion

The SSI project reveals several cultural factors that are important to consider in the design of communication systems for Deaf people. This project also provides design lessons regarding the use of dictated speech on shared tabletop displays.

4.6.1 Cultural Factors

ASL is the native and preferred language for many members of the Deaf community. There are cultural implications involved in designing an English-based technology for the Deaf. For general conversations, several participants preferred communication in ASL with an interpreter. Alex says that communicating through the interpreter allowed her to express herself best: "My identity is Deaf. I prefer interaction in ASL." The tradeoff happens when a conversation is extremely personal or when the Deaf individual wants independence from an interpreter. In this case, our Deaf participants indicated that communicating through the digital table provides enough benefit for them to use English instead of ASL. We are considering ways to make the technology more inclusive of Deaf users by including video transcription of sign language next to the English text.

Conducting research with a Deaf population presents specific challenges. For non-signing researchers, an interpreter must be onsite to help facilitate interviews and usability studies. We found that traditional means of recruiting, such as online postings and email, were inadequate. Involving community members and domain experts early in our process led to a partnership with our city's Deaf community services center. One of their members created a video blog posting in ASL that advertised our study. This was highly effective and illustrates the importance of reaching a population of interest through their preferred language and communication medium.

4.6.2 Dictated Speech on Digital Tables

Speech recognition is a promising technology for supporting natural forms of interaction around digital tables. Compared to previous work on spoken commands (Tse et al., 2008, 2006), dictated speech presents new interaction challenges for both Deaf and hearing populations. The following design guidelines increased interface usability and speech recognition:

1. The system should limit the impact of ambient noise, but the speaking user should not have to turn a microphone on and off. An on/off button adds an extra layer of unnecessary complexity and work for the user.

2. The speaking user should have control over the speech that is added to the display. Our system presents the user with three best guesses from the recognizer. This gives control to the speaker, allowing them to select only accurately detected words and phrases. While this design requires an additional step, it greatly improves accuracy.

3. The interface should enable conversational repair. Mistakes in recognition and conversation will happen, so it is important to provide users with a mechanism for repairing their speech. SSI enables this through "trashcans" on the interface, intentionally placed in the corners of the display to increase situation awareness by others.

4. The interface should provide an auxiliary way to enter speech. New words and technical names may stump the recognizer. In our design, we provide a virtual keyboard for the speaking user to enter unusual words and found that this alternative worked well.

5. Application designers should take into account the tolerance of their user population and pace of conversation in the domain of interest. The doctor in our study said that speech input was useful, but that some doctors would not have the time or patience for voice recognition software and might prefer a second physical keyboard.

4.7 Conclusion

This chapter presented the design and early analysis of SSI. We demonstrated that multimodal tabletop displays are a promising alternative for augmenting communication between Deaf and hearing people. SSI allows conversation participants to maintain face-to-face interaction while enabling privacy and independence. Research on SSI contributes to the growing interest in multimodal, multitouch cooperative systems and complements previous work in the field. The next chapter describes an in-depth analysis of interaction using SSI.

Chapter 4, in part, is a reprint of the material as it appears in the Proceedings of the 2008 ACM Computer-Supported Cooperative Work (CSCW). Piper, A.M. and Hollan, J. Supporting Medical Conversations between Deaf and Hearing Individuals with Tabletop Displays. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Analyzing Multimodal Interaction around a Surface Computer

This chapter examines interaction through Shared Speech Interface (SSI). SSI is an application for a multimodal tabletop display that supports communication between a hearing and Deaf individual by converting speech to text and representing dialogue history on a shared interactive display surface. We compare communication mediated by SSI with communication mediated by a human sign language interpreter. Results indicate that the multimodal tabletop display: (1) allows the Deaf patient to watch the doctor when she is speaking, (2) encourages the doctor to exploit multimodal communication such as co-occurring gesture-speech, and (3) provides shared access to persistent, collaboratively produced representations of conversation.

5.1 Introduction

Loss of hearing is a common problem that can result from a variety of factors (e.g., noise, aging, disease, and heredity). Approximately 28 million Americans have significant hearing loss, and of that group, almost six million are profoundly deaf (NIDCD, 2008). A primary form of communication within the United States Deaf community is American Sign Language (ASL). ASL interpreters play a central role in enabling face-to-face communication between Deaf and hearing people. For many Deaf people, communicating through an interpreter is an optimal choice. Interpreters, however, are expensive and unavailable in many situations. Furthermore, though interpreters are bound by a confidentiality agreement, the presence of a third person in a private conversation may reduce a Deaf person's comfort and inhibit their willingness to speak candidly. These factors are especially relevant for the topic of our current analysis: medical conversations between a Deaf patient and a hearing, non-signing doctor.

We designed and evaluated SSI, a multimodal tabletop application that facilitates communication between a Deaf and hearing individual. The application was designed to enable private and independent communication within the context of doctor-patient consultations. Findings reported in Chapter 4 indicate that communicating through a multimodal tabletop display is both feasible and desirable for Deaf individuals; however, it is not yet clear how the tabletop display affects communication on a cognitive and social level. Drawing on theories of distributed cognition (Hutchins, 1995; Norman, 1993; Salomon, 1993), this chapter presents a micro-analysis of interaction between Deaf and hearing individuals to begin to address questions regarding communication, coordination, and cognition. Our analysis examines speech, gesture, eye gaze, and device interaction involving the doctor, patient, and sign language interpreter. We find that the digital table provides dialogue with properties that are not available in conversation through a human interpreter. Specifically, the digital table transforms ephemeral dialogue into a lasting form that allows the Deaf individual to better attend to the speaker, supports co-occurring gesture-speech by the hearing user, and provides a shared visual record of conversation.

5.1.1 Deaf Communication

While not all Deaf or hearing impaired individuals use sign language, sources estimate that ASL is the fourth most widely used language in the United States (NIDCD, 2008). Sign language interpreters are often used to facilitate communication between Deaf and hearing individuals, but access to an interpreter requires foresight and can be expensive. While interpreter services are important, they raise issues of privacy in communication. The Deaf community in many locations is small and well-connected. It is not uncommon for a Deaf person to know the interpreter, which creates concern for very personal conversations. The interpreter scheduled on a given day may also be of the opposite gender, making discussion of certain medical issues even more uncomfortable. Face-to-face communication through an interpreter requires the Deaf individual to focus their attention on the interpreter rather than the speaker. Taking notes during conversation involving an interpreter is also challenging because the Deaf individual must pay close attention to the interpreter and cannot easily look down to make notes on paper. Not all Deaf individuals know how to read and write in a spoken language such as English, but those who are proficient may use hand written notes to communicate in the absence of an interpreter. Communication with hearing people is further complicated because sign languages are not simply visual forms of spoken languages. Instead, each sign language has its own unique grammatical and syntactical structure, making a spoken language a second language for many Deaf individuals.

Technology has transformed communication for the Deaf community. Telephone use was impossible for Deaf individuals until the adaptation of the Teletype machine (TTY) which allowed individual lines of keyboard entry to be transmitted over phone lines. Adoption of the TTY, its subsequent electronic versions, and now the personal computer, made typing an essential mode of communication within the Deaf community. Researchers have developed a variety of technologies to address communication barriers between the Deaf community and hearing world. As early as 1975, researchers began investigating how cooperative computing environments, such as early forms of Instant Messenger, could facilitate communication between Deaf and hearing individuals (Turoff, 1975). More recently, human-computer interaction researchers have examined how mobile devices (e.g., Cavender et al., 2006), tablet computers (Miller et al., 2007), and browser based technologies (Schull, 2006) can augment communication for Deaf individuals. While these solutions address various communication challenges for Deaf people, none address face-to-face communication around a single shared display.

5.1.2 Multimodal Tabletop Displays

Digitally enhanced tabletop displays are growing in availability (e.g., Microsoft Surface). The ability to receive multiple simultaneous touch inputs from a number of people makes tabletop displays a promising technology for facilitating face-to-face group interaction. Within the field of human-computer interaction, substantial attention is given to how tabletop displays can support face-to-face communication and mediate group social dynamics (see Morris, 2006, for a review). Compared to vertical displays, such as a computer monitor or wall mounted display, tabletop displays result in more equitable interaction and shared responsibility by group members (Rogers and Lindley, 2004). Recently, there has been growing interest in multimodal multitouch tabletop systems. A multimodal tabletop system integrates touch input along with speech, eye gaze, and/or stylus input. Tse and his colleagues (2007) explore how multimodal tabletop systems support gaming, pair interaction around a multimodal tabletop display, and techniques to wrap single-user applications so they include multimodal interaction. Researchers have examined a variety of tabletop group work issues with hearing populations, but untilh the Shared Speech Interface project, researchers had not examined tabletop computing scenarios with hearing impaired populations.

To address this gap, we developed Shared Speech Interface (SSI), a multimodal tabletop application that enables collocated face-to-face communication and cooperative activity between a hearing and Deaf individual. The design of SSI exploits the affordances of multimodal tabletop displays while addressing the communication needs for a Deaf patient and a hearing, non-signing medical doctor. Consultations with physicians often involve visuals, such as medical records, charts, and scan images. Interactive tabletop displays are effective for presenting visual information to multiple people at once without necessarily designating one person as the owner of the visual. Taking notes while meeting with a physician is problematic for Deaf individuals because it requires simultaneously attending to the doctor's facial expressions, the interpreter's visual representation of speech, and notes on paper. A multimodal tabletop display allows the doctor and patient to maintain face-to-face contact while viewing a shared, interactive representation of their conversation and other visual materials.

SSI runs on a MERL DiamondTouch table (Dietz and Leigh, 2001) and uses the DiamondSpin toolkit (Shen et al., 2004). The DiamondTouch table is a multiuser, multitouch top-projected tabletop display. People sit on conductive pads that enable the system to uniquely identify each user and where each user is touching the surface. SSI supports conversational input in English through standard keyboard entry and a headset microphone. Audio captured from the microphone is fed into a speech recognition engine, converted from speech to text, and then displayed on the tabletop interface. Currently, SSI works for two users communicating in a face-to-face setting. The hearing user speaks into the headset microphone and the Deaf individual enters speech through a standard peripheral keyboard. As the two individuals communicate, their speech appears on the tabletop display in the form of moveable speech bubbles. See Chapter 4 for a detailed description of the system design.

Figure 5.1: A medical doctor and a Deaf patient communicate using Shared Speech Interface.

5.2 Analysis of Multimodal Human Interaction

While a tabletop display is considered multimodal when it has multiple modalities of input (e.g., touch and speech, or touch and eye tracking), interaction with other people around a tabletop display is inherently multimodal (Garfinkel, 1967; Schegloff and Sacks, 1973; Schegloff, 1984; Goffman, 1981; Kendon, 1967, 1980; Goodwin, 1981, 1986). In this chapter, we use video analysis techniques to closely examine the interplay between speech, gesture, and eye gaze as well as interaction with the device. Video analysis is routinely used to understand activity within naturalistic settings (e.g., Heath, 1986), but some laboratory studies also include analysis of multimodal human interaction data (e.g., Bekker et al., 1995; Kraut et al., 2003; Kirk et al., 2005). From a methodological perspective, Kirk et al. (2005) note the importance of studying laboratory data in an "ethnographic fashion." Furthermore, Hollan et al. (2000) argue more directly for an integrated approach to human-computer interaction research based on theories of distributed cognition and a combination of ethnographic and experimental techniques.

5.2.1 Gesture in Collocated and Remote Interaction

There is a growing interest in collocated gestural interaction and its relevance to the design of cooperative computing systems. Tang (1991) noted the pervasive nature of hand gestures in a group drawing activity and indicated the need to better understand this activity in relation to the people and artifacts in a collocated workspace. Bekker et al. (1995) studied gestures as a way of informing the design of cooperative systems. Kraut et al. (2003) examined how visual information, especially deictic reference, enabled situational awareness and conversational grounding in face-to-face, video-based, and audio-based interaction.

The horizontal form factor of tables provides unique affordances for group work compared to vertically mounted displays. Work by Rogers and Lindley (2004) noted an increased use of gesture when groups interacted around a tabletop display compared to a whiteboard display. In another study, Rogers et al. (2004) found that touching a display with fingers has ancillary benefits for group work, such as supporting turn-taking. With respect to gesture, Tse et al. (2007b) provided similar observations of pairs interacting around a multimodal tabletop display. They noted that "speech and gesture commands serve double duty as both commands to the computer and as implicit communication to others."

Several systems examine how representing gesture and eye gaze across remote environments affects interaction (see Tang and Minneman, 1990, as an early example). Kirk et al. (2005) examined how specific hand gestures within the context of remote cooperative activity promote awareness and coordinate object focused actions. Similarly, Luff et al. (2006) examined how people working remotely use pointing gestures to coordinate and align themselves around objects of interest.

5.2.2 Gesture Analysis

The term *gesture* is polysemous for human-computer interaction researchers interested in touch-sensitive surfaces. On one hand, gestures are commands to a computer system administered by touching or moving an object, finger, or hand on an interactive surface. In a more traditional sense, the term gesture refers to the way in which people move or use their body as a means of communication or expression of oneself or with others. This section focuses on this latter meaning of gesture. Recently there has been a growing interest in using gesture analysis to understand communication between people (McNeill, 1992; Kendon and Muller, 2001) and within cooperative work environments (Goodwin and Goodwin, 1996; Hindmarsh and Heath, 2000; Zemel et al., 2008). This interest is largely driven by a theoretical shift from considering gesture as peripheral to human interaction to viewing gesture as central to communication and thought. Distributed cognition treats the body, and subsequent gestures we produce, as a central part of the cognitive activity system (Hutchins, 1995; Becvar et al., 2005). Kendon (1980) was one of the first to articulate the perspective that speech and gesture are inextricably linked. McNeill proposed a theory that speech and gesture involve a single conceptual source (McNeill, 1985, 1992). He posits that speech and gesture acts develop together. Goldin-Meadow (2003) suggests that gesture can indicate underlying reasoning processes that a speaker may not be able to articulate, and therefore, a better understanding of gesture promises to play a crucial role in teaching and learning (see Roth, 2001, for a review).

For the purposes of our discussion and in agreement with practices of gesture researchers, we identify gesture as spontaneous movements of body or hands that are often produced in time with speech but may also occur in the absence of verbal utterances (see McNeill, 1992). Actions such as head scratching or moving an object in an environment are not considered gestures. In our analysis, we pay particular attention to gestures that communicate and mediate activity. A descriptive analysis of gesture provides insight into how people exploit their bodies and environment during multimodal tabletop interaction.

Within gesture research, sign language is considered a separate class of communication. Each sign language has a specific syntactical and grammatical structure, and specific gestural forms within a sign language take on linguistic meaning. Communicating through sign language, however, does not preclude the use of spontaneous gestures. In fact, signers use the same proportion of meaningful gesture as speaking individuals use in verbal dialogue (Liddell and Metzger, 1998). There is growing evidence that people – both hearing and Deaf – attend to and interpret information in gestures (Goldin-Meadow, 2003; Cassell et al., 1999; Beattie and Shovelton, 1999).

5.2.3 Eye Gaze Analysis

In addition to gesture, other modalities of interaction, such as eye gaze, can provide insight into communication. Early work by Kendon (1967) gives a history of gaze research and describes the function of gaze as “an act of perception by which one interactant can monitor the behavior of another, and as an expressive sign and regulatory signal by which he may influence the behavior of the other.” Change in gaze direction such as looking away while speaking and then back to the listener at the end of an utterance gives listeners information about turn-taking (Duncan, 1972, 1974; Duncan and Fiske, 1977). Eye gaze is also used to demonstrate engagement (Goodwin, 2000, 1981) as well as indicate attention and show liking (Argyle and Cook, 1976; Kleinke, 1986) during face-to-face interaction. Eye gaze, accompanied with or without gesture, is also used in pointing acts (Kita, 2003).

When working with Deaf populations, understanding patterns of eye gaze is especially important. Direction of gaze indicates whether or not an individual is attending to visual forms of speech. In conversation, a Deaf individual reading sign will maintain relatively steady gaze towards the person signing (Baker and Padden, 1978; Siple, 1978). Eye contact with the signer is a signal that the signer has the floor, and shifting gaze away from the signer can indicate a turn request (Baker, 1977). In American Sign Language, the direction of gaze may be used for deictic reference (Baker and Padden, 1978; Engberg-Pedersen, 2003), and monitoring gaze direction may provide insight into the accompanying interaction. Signers tend to shift gaze from the face of their listener to their own hands when they want to call attention to gestures, and it is common for the signer to look back up at their listener to ensure that they too are looking at the gesture (Gullberg and Holmqvist, 2006). Work by Emmorey et al. (2008) found that people reading sign language do in fact follow gaze down to the hands when a signer looks at his or her hands. Eye gaze is an important aspect of human-human interaction, and understanding it may lead to deeper analyses of cooperative multimodal systems.

5.3 Experimental Setup

Eight Deaf adults (mean age=33, stdev=11.4, range=[22,52]; 3 males) and one medical doctor (age=28, female) participated in a laboratory study. All eight Deaf participants were born Deaf or became Deaf before the age of one. Three participants identified English as their native language and five identified ASL. All participants were fluent in ASL and proficient at reading and writing in English. The medical doctor had prior experience treating Deaf patients but did not know ASL. None of the participants had used a tabletop display prior to participating in this study.

Deaf participants were given sample medical issues (e.g., about routine vaccinations for travel abroad or advice on losing or gaining weight) to discuss with the doctor. Each Deaf participant worked with the same doctor, which resembles the real-world scenario where one doctor has similar conversations with multiple patients throughout the day. The patient and doctor discussed a medical issue using either the multimodal tabletop system (digital table condition) or a professional American Sign Language interpreter (interpreter condition)[1]. Each discussion prompt had a corresponding medical visual that was preloaded into the tabletop system (e.g., a map for discussion about foreign travel). A paper version of the visual was provided for the interpreter condition. Medical professionals helped to ensure that the discussion prompts reflected authentic conversations that might occur in normal patient interaction but whose content did not require participants to discuss information that might be too personal. Deaf participants experienced both the digital table and interpreter condition. The order of conditions and discussion prompts was randomized between subjects. Each session was video taped by two cameras from different angles to capture participants' interactions with each other and the digital table. All sessions were conducted around a DiamondTouch table to

[1]Chapter 4 describes a third Mixed condition that involves both the Digital Table and Interpreter. However, we focus analysis in this chapter on the Digital Table and Interpreter conditions to provide clearer comparisons of interaction.

keep the environment consistent; the tabletop display was turned off for the interpreter condition. Three researchers were present for the testing sessions and took notes. Each conversation with the doctor lasted between seven and nine minutes.

Our research team reviewed over 200 minutes of video data and transcribed and coded key segments of interaction. We were careful to select segments of activity that are representative of larger patterns in the data set. Video data were transcribed using notation techniques by Goodwin (2000) and Jefferson (2004). Brackets surround speech that is co-timed with a gesture, and bold-face speech indicates the *stroke* of the gesture. Transcriptions involving the interpreter indicate the interpreter's speech on behalf of the Deaf individual and are not a transcription of sign language.

5.4 Results

Initial findings indicate that Shared Speech Interface is a promising medium for facilitating medical conversations (see Chapter 4), but how does the multimodal tabletop display shape communication? To answer this question, our analysis focuses on four areas of collocated interaction. First, we examine patterns of gaze by the Deaf individual as a way to understand their attention during interaction. Second, we present an analysis of gesture by the doctor to identify differences in how she exploits multiple modes of communication depending on the communication medium. Then, we discuss how the Deaf individual monitors multiple modalities of communication with an emphasis on co-occurring gesture-speech by the doctor. Lastly, we describe how the tabletop display provides persistent, collaboratively produced representations that can aid discussion in cognitively valuable ways.

5.4.1 Use of Eye Gaze

Video data reveal distinctly different patterns of eye gaze by the Deaf individual when conversation is mediated by an interpreter compared to the multimodal digital table. Eye gaze is a particularly critical channel of communication for Deaf individuals, as conversation is purely visual. Examining eye gaze data allows us to infer where the Deaf individual is attending during communication. Our results show that when an interpreter is involved in communication, the Deaf individual focuses gaze on the interpreter and glances only momentarily at the doctor, as expected per Baker and Padden (1978) and Siple (1978). We found that Deaf participants in our study looked at the interpreter when they were reading signs (i.e., “listening”) as well as when they were signing (i.e., “speaking”). Consider the following excerpt of conversation from the interpreter condition. In this interaction, the doctor fixes her gaze on the Deaf patient; however, the Deaf patient focuses primarily on the interpreter and makes limited eye contact with the doctor. In both conditions, the doctor maintains eye contact with the patient throughout the conversation and uses eye gaze and backchannel communication (e.g., head nodding in center frame of Figure 5.2) to demonstrate attention and agreement with the patient’s speech.

((*patient fixes gaze on interpreter*))
((*doctor fixes gaze on patient*))

D: where do you plan on eating (.)
most of the time (3.2)

P: uh:: (.2) i well (.)
im not exactly sur::e (.2) ((*nods*))
but maybe (.) probably (.)
ill stay with the group that
were going with (.)

D: mmhumm (.5) mhum=((*nods*))

P: and eat in the jungle (.3)
id imagine=

P: ((*shifts gaze to doctor*))
=or im not sure (.3)
((*shifts gaze back to interpreter*))
theres no restaurants there in the jungle (.2)
so i dont know (.6)

Figure 5.2: Doctor and patient communicating through interpreter. Patient watches interpreter while doctor looks at patient.

To elaborate this point, consider Figure 5.3, which illustrates the duration and patterns of eye gaze by this same individual. We highlight this case because the pattern illustrated here is typical for interaction. In the interpreter condition, the patient fixes her gaze on the interpreter as needed for communication (Figure 5.3, grey areas in top bar graph). In contrast, communication through the digital table allows her to spend more time watching the doctor (Figure 5.3, black areas in bottom bar graph). As illustrated by Figure 5.3, when an interpreter mediates communication, this Deaf patient makes quick one-second glances at the doctor and rarely holds gaze for longer than 3 seconds (gaze time on doctor: total=77 sec, mean=2.1, stdev=2.0; gaze time on interpreter: total=293 sec, mean=8.0, stdev=7.3). This is likely an attempt to demonstrate that she is attending to the doctor without signaling to the interpreter that she would like a turn to speak, as a sustained shift in eye gaze in sign language communication indicates a turn request (Baker, 1977). In the digital table condition, the patient makes frequent shifts in gaze between the doctor and tabletop and looks at the doctor for slightly longer intervals (gaze time on doctor: total=143 sec, mean=3.0, stdev=2.6; gaze time on table: total=227 sec, mean=4.9, stdev=7.7). The digital table requires the patient to look down for periods of time to type speech on the keyboard. Even with looking down at the keyboard, the doctor in our study noticed a difference in eye gaze by the patient. In a follow-up interview she said:

> The physician-patient interaction involves more than just words. Body language is integral to the medical interview and provides key details into the patient's condition and level of understanding. The inclusion of the interpreter forced the Deaf patients to make eye contact with her rather than me, not allowing me to gauge whether information or a question I asked was understood as well as more subtle insights into the patient's overall comfort level.

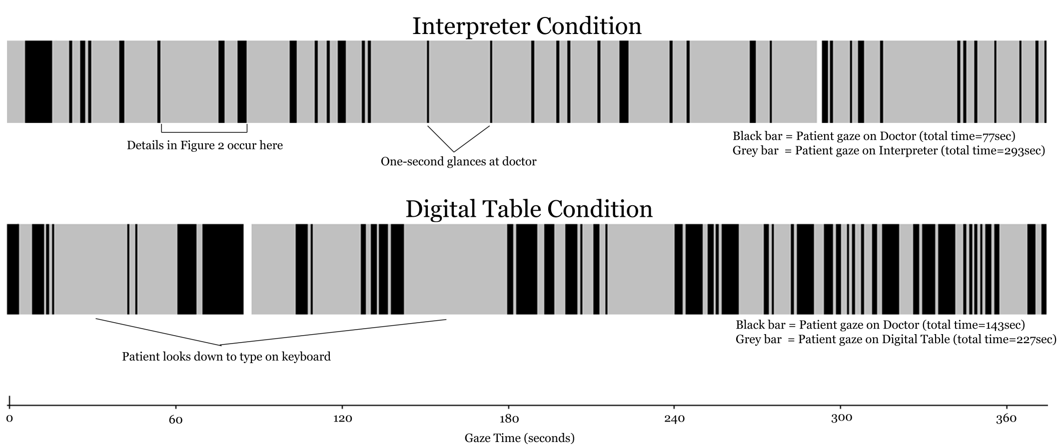


Figure 5.3: Duration and patterns of eye gaze by the Deaf patient during the Interpreter and Digital Table conditions

5.4.2 Use of Gesture

Communication through the digital table allows the patient to look at the doctor instead of requiring constant focus on the interpreter. Since speech appears in a permanent form on the tabletop display, the urgency of attending to the visual representation of talk is reduced. This allows both the doctor and patient to attend to and exploit multiple modalities of communication. Voice recognition capabilities free the doctor's hands and enable co-occurring gesture-speech in a way that traditional keyboard entry does not afford. Research on synchronized gesture-speech indicates that this activity is often co-expressive and non-redundant, and therefore provides interactants with multiple forms of information (McNeill, 1992). Consider another example of interaction in Figures 5.4. Here, the doctor recommends hand washing techniques to the Deaf patient by exploiting multiple modalities of communication including speech, gesture, and eye gaze. First, the patient looks at the doctor as she says "I would recommend." Then the doctor adds her speech to the display and continues "that you wash your hands." Both the doctor and patient look down at the display. Then the patient, likely to demonstrate understanding, holds up his hands and nods his head. The Deaf patient's action is an iconic gestural response to the doctor's speech (McNeill, 1992). As he gestures, he shifts his gaze from the tabletop to his hands, presumably to call the doctor's attention to his gesture (Gullberg and Holmqvist, 2006; Emmorey et al., 2008).

The patient then looks back at the doctor (Figure 5.4, middle row, left) as she formulates a recommendation for the patient. She makes a hand rubbing gesture as she says "with um." Then she uses the virtual keyboard to type the word "purell." The patient sees this word and responds by typing "Is that a specific brand soap?" His typing occurs simultaneously with the doctor's speech (Figure 5.4, middle row, right). The doctor's response (Figure 5.4, bottom) demonstrates that she attends to the patient's question for clarification. A critical moment in this interaction occurs in the bottom left image of Figure 5.4. The doctor and patient make eye contact as the doctor performs an iconic hand rubbing gesture timed with the words "alcohol based." Her gesture communicates the method of use for hand sanitizer, as alcohol-based sanitizers work by evaporating when rubbed into the hands. After this, both look down at the display to see the doctor's speech. Finally, the patient performs an emblematic OK gesture while nodding his head.

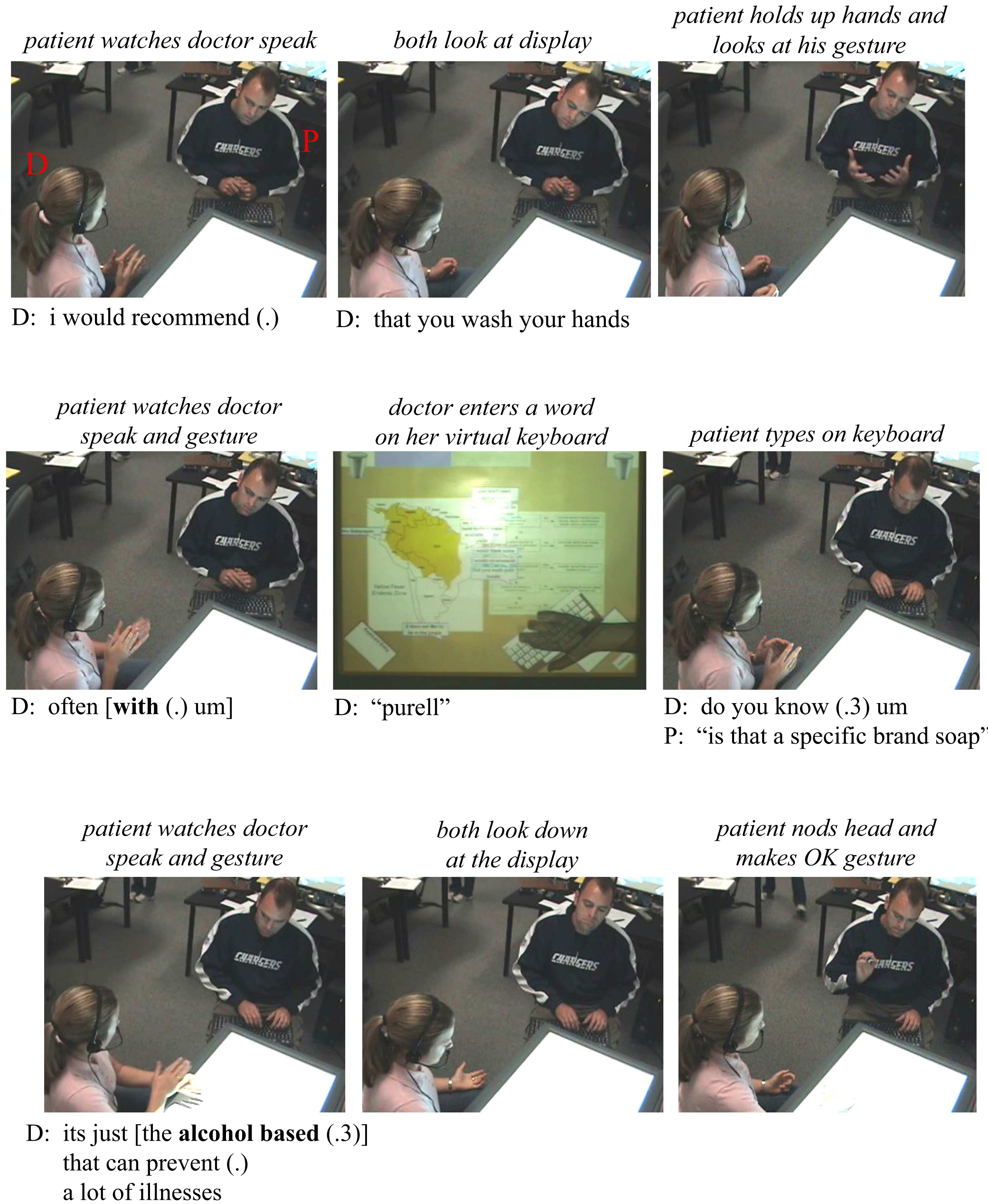


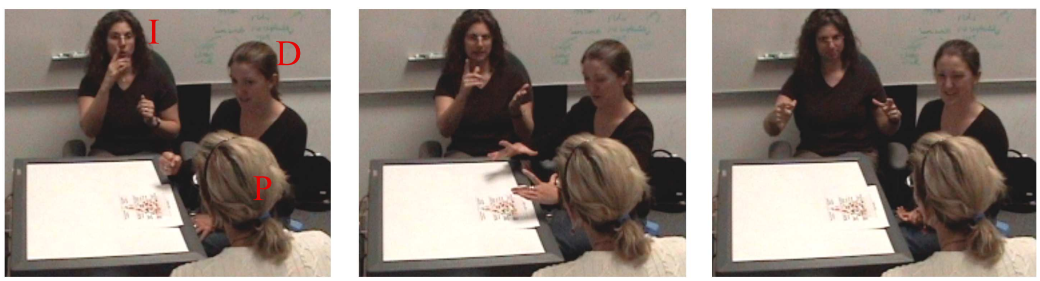
Figure 5.4: Doctor and patient communicate about hand washing through the digital table.

The doctor's carefully timed speech and gesture provide the patient with two pieces of information. First, her speech indicates the specific type of soap. Second, her gesture demonstrates how the soap is used. This information taken together yields a richer communicative form than either channel in isolation. This example demonstrates the importance of freeing the speaker's hands so that she is able to gesture as well as allowing the Deaf individual to attend to the speaker's gestures instead of maintaining focus on the interpreter. In this example, and in others, we were struck by the highly coordinated use of speech, gesture, and eye gaze between the doctor and patient. The doctor's rich use of gesture to augment speech occurred often in interaction through the digital table. Similar use of gesture was *not* observed when the interpreter was present.

In a follow-up interview, the doctor said that she intentionally tried not to gesture when the interpreter was present. She went on to explain that she did not want to compete with the interpreter for the patient's visual attention. In addition, interaction without the interpreter allowed the patient to frequently look at the doctor during communication, as is shown in Figure 5.3. This was a common pattern in the data. Having a larger percentage of the Deaf patient's visual attention may have encouraged the doctor to elaborate her explanations with gesture (although this hypothesis needs to be examined with additional studies). Our analysis suggests that the multimodal tabletop system allows the doctor and patient to attend closely to each other's use of speech, gesture, and eye gaze as mechanisms for mediating communication. This also enables the doctor and patient to better monitor and exploit multiple modalities of communication, such as co-occurring gesture-speech.

5.4.3 Monitoring Multiple Modalities of Communication

One challenge for Deaf individuals involves monitoring multiple sources of visual information during conversation. Noticing and attending to co-occurring gesture-speech is a particularly challenging process when communication is mediated by an interpreter. Interpreter-mediated communication requires the Deaf individual to notice co-occurring gesture-speech by the speaker and then put the speaker's gestures in context of the interpreter's gestural interpretation. Professionally trained interpreters are highly skilled, but they only occasionally replicate a speaker's gestures. Furthermore, through interviews with professional interpreters, we found that their formal training does not specify when, if ever, they should replicate gestures made by the speaker. Overall, there were limited speech-gesture acts by the doctor in the interpreter condition, but this behavior did happen occasionally. Figure 5.5 is an example of the doctor using co-occurring gesture-speech. Here, she makes a fist like gesture (left) and then a two-handed iconic gesture (middle) to clarify portion size. Timing of speech and gesture is an issue, as the doctor completes each gesture before the interpreter begins signing her speech. In this example, the interpreter did in fact recreate the doctor's gestures in context of her sign language interpretation but often the interpreter may not recreate the speaker's gesture, meaning that for at least a portion of communication the Deaf individual must notice and attend to the speaker's gesture on their own. Even in cases in which the interpreter does recreate the gesture, it may not be formed or timed in exactly the same way as the original, creating interpretation challenges. In contrast, communication through the digital table provides opportunity for the Deaf individual to look directly at the speaker's gestures, and as Figure 5.4 illustrates, gestures played an important role in establishing a shared understanding.



D: [**cooked pasta should only** be the size of your fist] (.4)

D: [not the **big bowls** (.)]

D: that are served in restaurants

Figure 5.5: Doctor uses gesture with her speech. Interpreter relays speech and gesture information.

5.4.4 Persistent, Collaboratively Produced Representations

Unlike other assistive technologies that enable communication between Deaf and hearing individuals, the shared tabletop display provides a central focal point and space for establishing common ground (Clark and Brennan, 1991). The horizontal tabletop surface provides a space through which the doctor and patient cooperatively create, view, and manipulate representations of conversation. The shared conversation space allows the doctor and patient to gesture around and point to previous speech, thereby anchoring their gestures to objects (physical and virtual) in the environment (Clark, 2003). Referencing interface objects most often occurs through situated, context-specific pointing gestures (Goodwin, 2003b). Both hearing and Deaf participants used deixis to reference the material and symbolic world in front of them. With the interpreter, there is no record or explicit external representation of discourse. Consider Figure 5.6 (top row) where the doctor annotates a food pyramid diagram. Here, the doctor uses pointing gestures on a food pyramid diagram as she explains a balanced diet. The Deaf patient must attend to both the interpreter's interpretation of speech as well as the doctor's pointing gesture occurring with her speech.

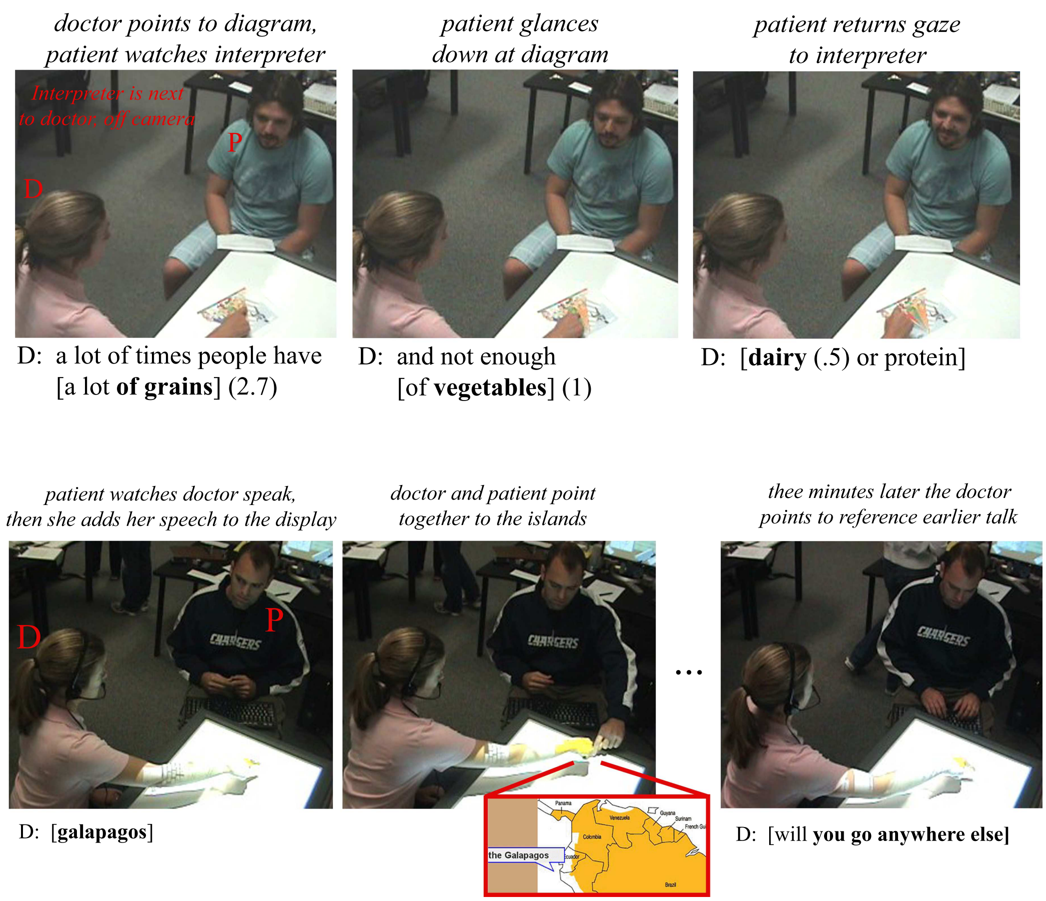


Figure 5.6: Top: Doctor points to parts of a diagram as she speaks. Patient monitors interpreter and doctor's pointing gestures. Bottom: Using the digital table, the Doctor labels the Galapagos Islands on the map and then points to the speech bubble three minutes later.

In this example, the doctor uses her speech and pointing gestures to walk the patient through parts of a food pyramid. Each time she points to a section of the diagram, she shifts her gaze to the table, likely an attempt to draw her listener's attention to the diagram. Several minutes later the doctor references this diagram again to summarize her recommendation about a well-balanced diet, but the conversation and gestures she made to the patient are now only a memory.

The digital table stores collaboratively created representations of speech and allows users to rearrange and manipulate speech bubbles. Images in the top row of Figure 5.6 illustrate challenges with pointing to parts of a diagram while speaking; the digital table uniquely supports this form of interaction. We observed an interesting form of pointing that occurred through the strategic placement of speech bubbles. The tangible and persistent nature of speech bubbles affords certain interactions by serving as manipulatable cognitive artifacts (Hutchins, 1995). A speech bubble gains meaning beyond its literal text depending on how it is situated, or anchored, with respect to other parts of the activity. The canonical shape of speech bubbles, specifically the tail, allows the doctor and patient to use the objects as a pointing mechanism. That is, participants strategically placed speech bubbles around the display so that the tail of the speech bubble touched a relevant part of the background or another speech bubble. Figure 5.6 (bottom center frame) provides an example of this behavior. In this interaction, the doctor uses a speech bubble to label and reference part of a map. The patient mentions that he is traveling to the Galapagos Islands. The doctor says "Galapagos" as she points, and the patient points along with her to clarify the location. Subsequently, the doctor moves the "Galapagos" speech bubble to label the islands on the map. Then she uses this action to show that the islands are outside the Yellow Fever endemic zone (bottom center frame of Figure 5.6) and explain that the patient will not need the Yellow Fever vaccine. Conversation continues, and the topic changes. Approximately three minutes later the doctor comes back to "the Galapagos" speech bubble. She points to the speech bubble while asking, "Will you go anywhere else?"

The persistent nature of speech along with the shared context of the tabletop display affords referencing both new and previously created external representations of speech. The persistent nature of speech also allows participants to review their entire conversation. Both the doctor and patients looked back over their previous conversation at some point during the activity. In a post-session interview, the doctor said, “It was good to look back at what I had covered with that particular patient,” and explained that, “[The digital table] would be helpful because it is not uncommon in medicine to have very similar conversations with different patients throughout the day.”

5.5 Discussion

Our analysis highlights differences in interaction between a Deaf and hearing individual when communication is mediated by a multimodal tabletop display as compared to a human sign language interpreter. These differences reveal several tradeoffs. Although speech recognition technology can not yet provide the richness and accuracy associated with translation by a competent interpreter, it does allows the doctor to exploit gesture for communicative purposes without fearing that she might distract the Deaf individual from the interpreter. One example is the hand washing iconic display coinciding with speech depicted in Figure 5.4. In addition, transcribed speech-to-text allows the doctor and patient to have a shared record of conversation. This provides new artifacts in the environment, enabling pointing and other gestures (Roth and Lawless, 2002). Removing the time-critical visual demands of interpreter-mediated communication allows the Deaf individual to focus more on the doctor while she is speaking. In turn, this helps the patient attend to the doctor’s speech-gesture acts and enables the doctor to better gauge patient understanding through increased eye contact. Speed of communication is another important tradeoff issue. Current speech recognition is no match for a skilled interpreter. When using the speech recognition system, the doctor must speak slowly and carefully to ensure accurate recognition. Another delay occurs when selecting the appropriate alternative from the output of the recognition system and in correcting it when required. However, necessitating slower dialogue on the part of the doctor is not an entirely negative outcome. Considering that English is a second language for many Deaf individuals, slowing the doctor’s speech could in fact be a positive cognitive consequence of communicating through the tabletop display.

SSI has the potential to benefit multiple user groups and enable new cooperative computing applications. Shared displays, especially tabletop displays, are beneficial for a variety of group work tasks. Since the inception of our project and the idea to visually represent conversation on a tabletop display, members of the Deaf community have mentioned numerous contexts in which this could be useful. Of these, the most frequently identified are counseling or therapy sessions, banking and financial services, meetings with an architect or interior designer, ASL and lip reading education, classroom group work, and retail environments. Beyond the Deaf community, the cognitive affordances of SSI have implications for individuals with moderate hearing loss as well as unimpaired hearing users. The challenge of medical conversations is certainly not restricted to the Deaf community. Because of associated stress and other factors, it is easy to forget details to tell the doctor and even easier to forget specific instructions given during consultation. The affordances of SSI, such as preloading questions for the doctor and referencing a transcript of a previous conversation, extend to all populations. Similarly, the ability to archive and subsequently revisit past multimodal conversations and collaborations has interesting potential to augment interaction.

The concepts behind SSI also have specific implications for user populations with other language-related communication barriers. For example, representing speech on a shared display has pedagogical benefits for language learning. Consider a case in which speech bubbles store textual and auditory information from a native speaking teacher and a student learning a second language. In this case, both textual and auditory representations can be accessed in a shared collaborative context. The availability of visual and spatial representations of language also stand to benefit individuals with linguistic processing disabilities, such as aphasia or apraxia of speech. Language could take on a variety of representations including textual, auditory, and pictorial forms. For these individuals and other populations, a shared, collocated workspace has considerable promise to help establish common ground and assist communication.

5.6 Conclusions and Future Work

Analysis of multimodal human interaction data is primarily used in ethnographic approaches to understanding everyday activity (e.g., Goodwin, 1981; Heath, 1986), but there is a growing interest in using multimodal analysis to understand interaction in experimental cooperative work settings (Bekker et al., 1995; Kraut et al., 2003; Kirk et al., 2005). We suggest that multimodal analysis can aid laboratory evaluations of tabletop technology as well as other cooperative work technologies in the following ways:

1. Analysis of eye gaze provides a metric for understanding how people coordinate visual attention.

2. Analysis of gesture and body movement illustrates the complexities of communication and participants' resourcefulness in various conditions.

3. The interplay between speech, gesture, and eye gaze can reveal cognitive and social consequences of new interactive media that would be difficult to detect with other methods.

Multimodal analysis, however, is tedious and extremely time-consuming. When analysis is so difficult, few analyses can be done and datasets are severely underutilized. Researchers come to have a large investment in the chosen data segments. Since each analysis may appear as an isolated case study, it can be difficult to know how common the observed phenomena may be. Larger patterns and contradictory cases can easily go unnoticed. Well-known human confirmation biases can affect the quality of the science when each analysis requires so much effort. The analyses presented in this chapter, for example, resulted from a year-long iterative process of analysis of video and audio data to understand how differing communication media shapes interaction. This form of detailed analysis plays an increasingly central role in our ongoing investigation of tabletop display systems.

In this chapter, we presented a deeper analysis of communication through SSI. We compared communication mediated by a multimodal tabletop display and by a human sign language interpreter. Results indicate that the multimodal tabletop display:

(1) allows the Deaf patient to watch the doctor when she is speaking, (2) encourages the doctor to exploit multimodal communication, such as co-occurring gesture-speech, and (3) provides shared access to persistent, collaboratively produced representations of conversation.

Chapter 5, in part, is a reprint of the material as it appears in the Proceedings of the 2009 European Conference on Computer-Supported Cooperative Work (ECSCW). Piper, A.M. and Hollan, J. Analyzing Multimodal Communication around a Shared Tabletop Display. The dissertation author was the primary investigator and author of this paper.

Chapter 6

Paper-Digital Interfaces for Aphasia Therapy

This chapter examines the practice of speech-language therapy for adults with aphasia. Aphasia is characterized by a reduced ability to understand and/or generate speech and language. Speech-language therapy helps adults with aphasia regain reading, writing, speaking, and auditory comprehension abilities. The therapy process is largely paper-based, making hybrid paper-digital interfaces a promising tool for supporting this practice. In this chapter, we present early findings from a year-long research project involving multimodal paper-digital interfaces to support speech-language therapy and communication involving adults with aphasia. We describe and analyze therapists' initial reactions to paper-digital interfaces and present two case studies of early prototype use by older adults undergoing speech-language therapy.

6.1 Introduction

According to the National Stroke Association, stroke is the third leading cause of death in America and a leading cause of adult disability (National Stroke Association, 2010). A stroke or cerebrovascular accident (CVA) occurs when an artery of the brain becomes clogged or ruptured and disrupts the blood flow to part of the brain. Brain tissue can die or be injured causing a change in neurological function. Strokes occurring on the left side of the brain often result in the loss of speech and language, commonly diagnosed as aphasia. The National Aphasia Association reports that of the estimated 400,000 strokes which occur each year, 80,000 result in aphasia (The National Aphasia Association, 2011). These individuals may understand written and spoken language but be unable to speak fluently or write. A stroke on the left side of the brain may also impair right side motor movement, including arm, hand, and leg functioning.

Other speech and language challenges may become present due to a stroke, traumatic brain injury (TBI), or neurological condition. For example, aphasia is often accompanied by apraxia of speech, where an individual is unable to create the voluntary mouth and tongue movements necessary for pronouncing words correctly. In addition, individuals may be unable to translate sounds into meaningful language or separate speech from background noise, also called auditory overload. Slurred speech (dysarthria) and swallowing problems (dysphagia) are other common challenges for this population.

The process of speech-language therapy is designed to help individuals with communication challenges regain their ability to speak, read, write, and understand spoken language. This chapter examines speech-language therapy, explores how digital pen and paper technology might support this practice, and introduces a multimodal paper-digital system into this setting.

6.2 Related Work

A range of augmentative and alternative communication (AAC) devices exist for supporting individuals during speech-language therapy and in their daily lives. Technologies may be augmentative (support verbal speech and language) and/or alternative (take the place of verbal speech and language). The majority of AAC devices provide multimodal representations of language through written text, speech or sounds, and images. Popular commercial devices include Dynavox (2011) and Lingraphica (2011), both of which are high-tech, picture-based systems with speech generators.

Research-based prototypes also support the needs of people with communication challenges, primarily individuals with aphasia. PhotoTalk, for example, is designed to help individuals with aphasia capture and manage digital photos that facilitate face-to-face communication (Allen et al., 2008). Moffatt et al. (2004) designed a sound- and image-enhanced daily planner to support individuals with aphasia. Chandler et al. (2009) created a mobile web application to support the process of word finding (or word searches) that individuals with aphasia often experience. In addition to helping individuals communicate more effectively and efficiently, some systems are designed to help users perform a task, such as cooking (Tee et al., 2005) or browsing the Internet (Devlin and Unthank, 2006), more independently.

The vast majority of HCI research has focused on purely digital AAC devices, yet this neglects the importance of paper-based practices. Myriad free or inexpensive paper-based resources are available to support speech-language therapy. These include, for example, paper activity worksheets, picture books, and communication boards that allow clients to point to various icons or text as a means of communication. It has been demonstrated in numerous settings (e.g., office work) that paper is a flexible medium with many affordances not supported by current digital systems (Sellen and Harper, 2001). Paper-based documents have important qualities for interaction that should be considered in conjunction with the benefits of digital communication tools.

In this work, we characterize the affordances of pen and paper interfaces in the setting of speech-language therapy, highlighting their importance and pervasiveness. We analyze how paper-based interfaces are currently combined with speech-based interactions. Based on this, we introduce a prototype that bridges the familiarity of current paper-based, low-tech AAC techniques with the dynamic nature of high-tech computer devices. We adapt a Livescribe Pulse Smart Pen (2011) to function as an AAC device to support therapy practices and communication. Finally, we describe how this hybrid paper-digital system stands to benefit speech-language therapy and older adults with aphasia.

6.3 Field Research

The first phase of this year-long project involves field observations and prototype exploration with 15 speech-language therapists (three licensed therapists and 12 therapy students) working with adults during structured speech-language therapy sessions.

Field research involved observations of interactions between therapists and clients, during which we recorded detailed field notes about therapy activities, face-to-face interaction, and the material environment. We also conducted regular interviews and prototype reviews. We video recorded therapy sessions and took still photos of the therapy environment. After each session, the therapist debriefed the researcher(s) for five to 30 minutes. This provided a chance to better understand why certain therapy techniques were used, any additional challenges the client or therapist faced that day, and to answer questions about observations. While speech-language therapists have clients with a range of conditions, such as dementia, Lou Gehrig's Disease (ALS), or Parkinson's Disease, our current work focuses on therapy practices with adults who have aphasia and apraxia of speech.

Theories of distributed cognition informed how we conducted field research and data analysis (Hutchins, 1995). The goal was to understand the intricacies of this interactive therapy practice and the challenges individuals with aphasia face in communication. As part of this goal, we attended to details that involved multiple modalities of interaction among people, the setup of the material workspace, and the evolution of practices over the course of a year. A person with aphasia may be able to speak only three words; however, this does not preclude an individual from acting as a competent speaker (Goodwin, 1995, 2003a, 2004, 2006). Goodwin's analysis of how his father with aphasia orchestrates the world around him to act as a competent speaker provides a framework for how we understood interaction involving individuals with aphasia.

6.3.1 Settings

We observed the practice of speech-language therapy in two settings: (1) the health center of a continuous care retirement community, and (2) a university-affiliated speech and language clinic.

Retirement community health center. The first setting involves ongoing observation of one speech-language therapist (JB) at a continuous care retirement community. Our research team observed JB on a weekly basis for four months, each week conducting one to three hours of observation. JB has been a licensed and practicing speech-language therapist for four years. He works at the retirement community and at other local senior care facilities. In the retirement community, we observed how the therapist worked with two older adult clients (age 88 and 90) during individual weekly hour-long therapy sessions. We detail the communication abilities and therapy needs of these two clients in Section 6.6, both of whom had a stroke resulting in expressive aphasia and apraxia of speech. Observations at the retirement community allowed our research team to understand the practices of one therapist over time and the progression of therapy techniques as client needs change.

University speech and language clinic. The second setting involved 12 hours of observation at a university-based speech and language clinic that provides services to adults and children with a range of communication challenges. The clinic is part of a graduate program that trains speech-language therapists and provides students with clinical experience. In this setting, we observed 12 therapy students working with clients (age 29-59) with aphasia and apraxia of speech resulting from a stroke or traumatic brain injury (TBI). One therapy student worked with each client during one-hour sessions. Some therapy students worked with their clients in a group setting while others engaged in one-on-one structured sessions. Family members and friends were able to observe the therapy sessions from the other side of a one-way mirror. After watching a session, the son of a man undergoing therapy said, "The biggest thing is that he's frustrated. He can understand everything but can't speak." In contrast to long-term observations at the retirement community, this environment exposed the research team to the broad range of communication needs of clients and therapy techniques employed.

6.3.2 Observational Themes

The health condition and therapy needs of any individual client are unique, yet field observations reveal several commonalities of speech-language therapy practices. Figure 6.1 illustrates two common scenarios for therapy, including the environment setup, orientation of participants, and materials. Below we summarize key observations from field research that inform the design of a paper-digital AAC system.

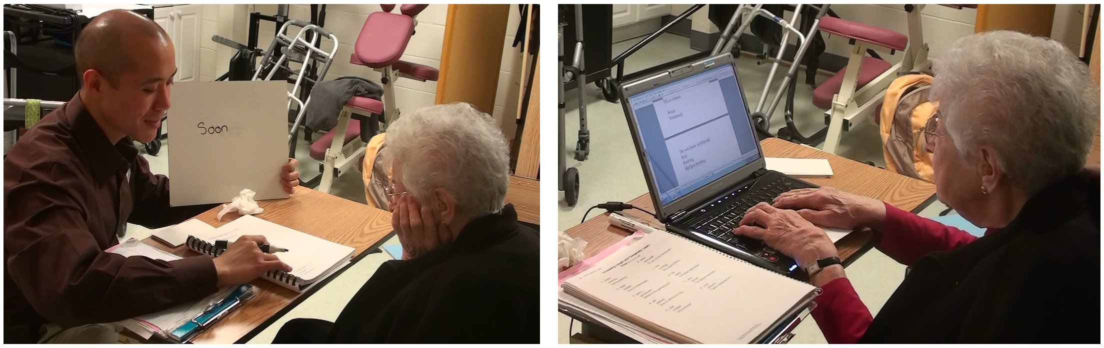


Figure 6.1: Common scenarios for therapy. *Left:* completing sentences using a paper workbook and dry-erase board. *Right:* transcribing sentences from a paper worksheet onto a computer. Images are of JB (therapist) and AF (client). All participants consented to the use of photographs for publication.

Speech-language therapy is collaborative and social. The process of speech-language therapy most often involves a therapist working face-to-face with a client at a shared desk or table. Both the therapist and client work collaboratively to reach a mutual understanding during conversation, each refining what the other has said until they reach shared agreement. For clients with a severe loss of speech or language, it may take the therapist and client 30 minutes to achieve joint understanding on a single point (e.g., if the client tries to explain what they did over the weekend). On some occasions, an individual undergoing therapy will practice language activities on their own as homework, but the majority of structured therapy interaction occurs in a social situation involving two or more people.

The therapy process is multimodal. Representations in multiple modalities reinforce and reaffirm spoken, written, and auditory forms of language. Figure 6.2 illustrates multimodal, collaborative interaction between one therapist and client. In this example, the therapist (JB) and client (AF) are both pointing to the word “join” in the workbook. The client reads this word as “jane” so the therapist says “join” and exaggerates his lip movements to emphasize the “o” sound. Therapists present clients with language information in various modalities depending on their communication needs.

Therapy is an ongoing process. Most therapists meet with their clients on a weekly or bi-weekly basis. Therapy is ongoing and builds on activities from the prior session.

Figure 6.2: Therapist and client sit face-to-face working on word pronunciation from a paper workbook. Therapist exaggerates as he pronounces the word "join" after the client mistakenly says "jane."

Therapists spend a few minutes at the beginning of each session reminding clients about what they did the previous session and reviewing homework. Since therapy is ongoing, therapists must manage plans and materials for clients over extended periods of time. They often organize paper-based plans in folders and notebooks so that they can document progress.

Sessions are tailored to each client. Even though several clients may have a similar diagnosis and present in a similar way, therapists must customize activities, goals, and materials for each client. Therapists consider the communication challenges and goals of each client and their individual therapy plan when generating activities for each session. Clients also progress at different speeds and require different forms of stimulation. Furthermore, therapists must be flexible as they address the client's needs from week to week. An emotional change in a client, for example, may render previously successful techniques unproductive.

Practices are rooted in pen and paper interaction. The structure and nature of speech-language therapy is largely influenced by existing practices with pen and paper. Regardless of a client's condition or communication goals, therapy materials are predominantly paper-based. Therapists use paper workbooks, worksheets, communications books, note cards, notepads, and books or novels during each therapy session. Clients whose ability to communicate is severely impaired may write words or draw pictures on blank pieces of paper. Higher functioning individuals may perform paper-based activities, such as reading from a word list, reciting minimal pairs (words that sound alike), or writing words to complete sentences. Many therapists also use paper to document treatment plans and notes about a client's progress.

6.4 Paper-Digital Prototype

Paper is an important medium for collaboration and shared interaction, and it is pervasive in speech-language therapy. Many affordances of paper play an important role in this setting: paper is inexpensive, flexible, portable, and disposable. Paper materials may be left with a client for home use or filed away by a therapist for documentation. Compared to other high-tech alternatives, pen and paper interaction is familiar and comfortable for most older adults, the predominant population undergoing speech-language therapy.

Existing technology for digital pen and paper (e.g., Anoto, 2011) enables static paper documents to be turned into interactive interfaces that can be used to recognize handwritten information, link to digital media, such as sound or voice recordings, or trigger specific digital applications on a separate computer either in real time (by streaming the collected information over bluetooth) or in batch mode (by connecting the digital pen to a computing device at a later time). This technology lets users exploit rich digital services while keeping the natural interaction common in traditional pen and paper interfaces. The system is based on a special digital pen that integrates a processor, memory, and an infrared camera able to interpret a unique dot-pattern printed on standard paper. By decoding the printed pattern, the device can track the pen's position on paper in real-time.

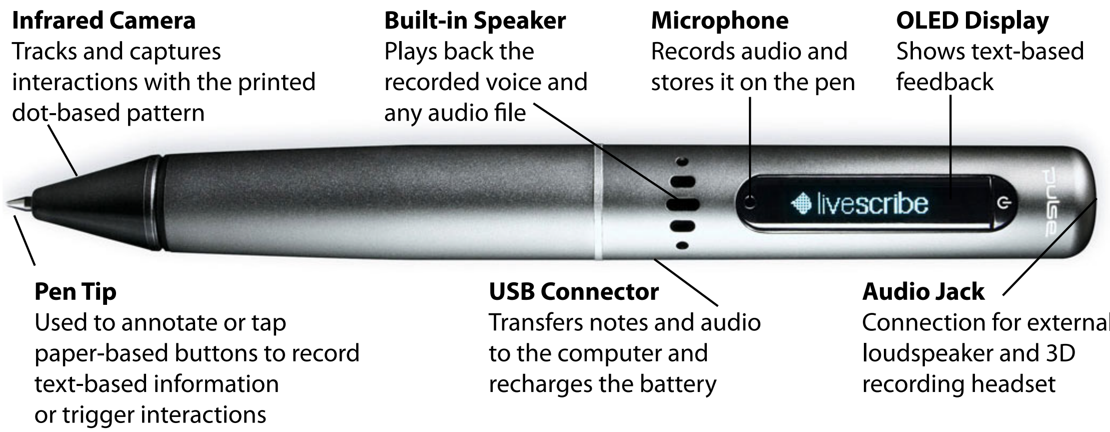


Figure 6.3: Livescribe Pulse Smart Pen with an integrated speaker and microphone for audio recording and playback.

Digital pen and paper technology has been explored for interaction with a range of digital documents and resources, including text processors (Weibel et al., 2008), PowerPoint presentations (Signer and Norrie, 2007), field scientist notebooks (Yeh et al., 2006), and collaborative annotations (Steimle et al., 2009). The technology, however, has yet to be explored as an assistive device for adults with communication challenges. The recent introduction of a novel kind of digital pen, the Livescribe Pulse Smartpen (see Figure 6.3), is particularly promising for speech-language therapy. This pen is a

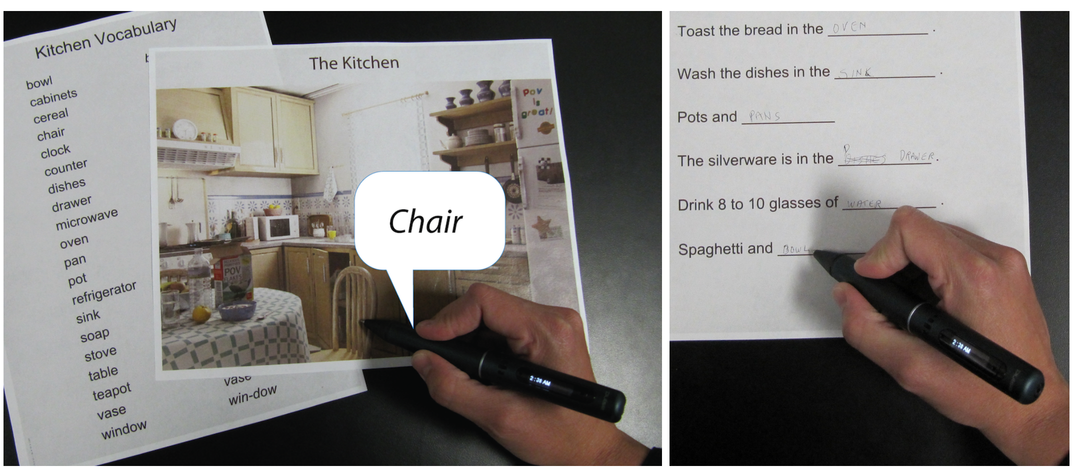


Figure 6.4: *Left:* Point-and-playback application, touching on an image or word plays associated audio. *Right:* Handwriting recognition, handwritten text is stored on the pen and played back.

"pentop computer" supporting multimodal interactions based on the combination of capturing and recognizing paper-based handwriting with audio recording and playback. In addition to these multimodal capabilities, Livescribe technology enables custom-made applications to be deployed directly on the pen. Field observations indicate that the combination of audio input and output with paper-based interactions may be useful for supporting speech-language therapy. As described earlier, pen and paper interaction is pervasive in current therapy practices, and both multimodality and collaboration are essential components of therapy. Moreover, Livescribe digital pens are relatively inexpensive (under $100) and readily available.

6.4.1 Design and Functionality

In order to design, develop, and deploy our custom applications and related interactive paper documents, we exploited the standard Livescribe development toolkit. This tool is an Eclipse plug-in requiring development in Java Micro Edition (J2ME) and is not appropriate for direct use by therapists unless they have programming experience. We developed sample interactive paper prototypes to demonstrate the Livescribe pen's capabilities to therapists. As shown in Figure 6.4, we produced interactive worksheets by modifying existing paper-based documents used during current therapy sessions. Interactive paper documents cover topics such as naming items in a visual scene display, completing sentences, speaking minimal pair words, listening to the name of a word or icon spoken aloud, and recording one's own speech. In addition, clients may practice writing words and phrases. The pen's handwriting recognition facilities automatically detect keywords and phrases that are written by hand. Practicing speech articulation and constructing sentences verbally is enabled through the pen's audio recording facilities. Clients may record phrases of speech and then play them back for reflection. Central to therapy practices, the system supports multimodal interaction and multimodal representations of language.

6.5 Prototype Review

As an initial step in the evaluation process, we reviewed the pen and paper prototypes with 15 speech-language therapists, including: JB at the retirement community, the two supervisors of the university clinic, and 12 therapy students. The prototype reviews were conducted in a focus group style to encourage idea generation and open discussion about the positive and negative aspects of this approach.

Appeal of pen and paper interaction. Therapists thought pen and paper was an appropriate medium for supporting interaction, especially for older clients. The simplicity of pen and paper interaction was key. For one man with aphasia and severe apraxia of speech, a clinic supervisor said, "This would be a great tool for him for a lot of reasons. For one, he's more of a paper and pencil kind of guy, which you saw today, he wanted to draw, draw, draw. He's been resistive to any kind of a device. We've talked to him about devices, we've talked to him about the Lingraphica." The clinic supervisor commented on another client, "[He] has had a lingraphica for over a year, but he hasn't touched it in over nine months... It's hard to navigate through, it's overwhelming. It's too much for a lot of clients." The other clinic supervisor said, "A lot of our clients don't want to use communication books or computers. They may be more inclined to use it if it's still a communication book model but all they have to do is point and it speaks." The simple, familiar, and straight-forward interaction of paper and pen was appealing to therapists given the cognitive and motor challenges many clients face. JB and the other therapists also wanted marking and non-marking pen tips (which Livescribe provides) and to laminate certain pieces of paper to make them hold up longer. The Livescribe pen works with laminated paper and paper behind a slip cover within a binder. Generating ideas for paper-based materials was not a challenge for therapists. A clinic supervisor suggested for one client, "We could have scanned the picture of his children on this dot paper and have the names right there because he obviously knows their names but just couldn't say them."

May reduce communication barriers. Several therapists thought that the system would allow them to better understand what a client "knows" by reducing the communication barrier. The student working with the man with severe apraxia said, “Even for [him], with something like that [the digital pen] we’d understand more of what he knows rather than assuming or estimating. I think he gets it, but having something more definite would be good.” Another student agreed, “It would give you insight into what they really know... Get past that communication barrier.”

Custom, human-generated audio. The ability for the pen to record and replay audio was important to therapists. JB wanted to use the recording facilities to help his clients reflect on how they said certain words. A student in the clinic said, “One of my clients I give multisyllabic words to practice at home for apraxia, but he doesn’t have the model of how to say it right, so if he at least had [the digital pen]... it would be good.” Many AAC devices use a computer synthesized voice, which may not provide clients with an accurate model of how to say words. Another student commented, “The pen puts intonation in there, which is different from a computer voice which doesn’t represent it right.” She continued, “My client has problems with multisyllabic words. He puts the intonation on the wrong syllable.” Recording custom, human-generated audio was an important feature of the system.

Enable independent practice for clients who live alone. JB explained that one challenge his clients face is that they often live alone and can become socially isolated. Residents may not have people with whom they regularly interact and practice communicating. A similar sentiment was echoed by therapists at the university clinic. One clinic supervisor said, “So many of our clients live alone... I’m thinking home programs for a lot of our clients who live alone, and they don’t have a partner or caregiver or someone to help them with their homework. They end up getting frustrated and not doing it.” While therapy is a collaborative and social process, supporting independent language activities at home is a positive benefit of the system. The clinic supervisor said to a student, “With your other client...he has this breakdown with phoneme representation... Again this is a gentleman who lives alone, so to have some kind of recording so he can hear. So he can write it and hear it, that would be really good for him.”

Time and cost. Therapists have limited time to prepare for each client, so naturally they were concerned about how much time it would take to prepare the interactive worksheets. Developing a way to help therapists quickly and easily create custom interactive worksheets is a main goal of our future work. For the present study, a researcher was available to generate custom content for therapists using the digital pen with clients.

In addition to time, therapists were concerned about the cost of the device, as this is a central challenge in adopting high-tech AAC devices. When told that the Livescribe pen costs under $200, the clinic supervisor said, “I think $200 is great, especially compared to an $8000 Lingraphica... The pen is very, very affordable, but the software and printer may be a concern.” Investing in a digital pen and paper system would require a retirement community or clinic to purchase a high-resolution color printer capable of printing the custom dot pattern. Such a printer costs around $1500 and could be a shared resource. Alternatively, a commercial printing service with only a per page charge could be used.

6.6 Case Study Evaluations

In addition to understanding the perspective of speech-language therapists, we wanted to explore how the pen and paper system might support a therapist working with older adult clients. We created custom therapy materials to support JB working with two clients at the retirement community: AF and MW. JB integrated the digital pen and interactive paper materials into his therapy plans for these clients. With each client, JB introduced the pen and paper system, demonstrated the desired interaction, and then supported the client as they learned to use the device. We examined how these two older adults learned to use, responded to, and adapted to the digital pen and paper system as well as JB’s impression of the device. To counteract background noise in the retirement community’s outpatient therapy room, we provided an external speaker to amplify the pen’s volume; however, the pen’s integrated speaker is sufficient in other environments such as the university clinic.

6.6.1 Case: Language Rehabilitation

AF (age 88, female) had a stroke in the left side of her brain approximately one year before we began field observations. As a result, she has aphasia and apraxia of speech that make it challenging for her to construct complete or coherent sentences and articulate words. However, AF is fairly high functioning in that she is able to say many words and is generally able to communicate her needs. AF lives alone but is well known throughout the retirement community. She often stops to greet other residents and asks "How are you, dear?" Her communication challenges become more evident during structured tasks where she needs to follow instructions or process competing auditory stimuli. She has fully recovered her motor ability in her hands although she uses a walker for balance.

Materials. Each week during therapy, JB provides a range of paper-based activities for AF. They use workbooks, worksheets, note cards, and any books AF is currently reading to practice word articulation and sentence construction. Before her stroke, AF was an excellent and avid cook. Given this, we created a set of interactive worksheets for AF that center around kitchen activities (Fig. 6.4, left). The worksheets include object identification with a visual scene display of a kitchen, a vocabulary list, sentence completion, following instructions, and written and auditory comprehension.

Initial use. JB introduced the digital pen and paper to AF, and she figured out the interaction within a couple of minutes. Because of her familiarity with one researcher, AF quickly recognized that the pen's voice was that of the researcher. The external speaker was necessary to provide sufficient volume for AF in the therapy room. While the pen was tethered to the speaker, the cord did not get in the way during interaction. However, the therapist had to monitor the connection and ensure that the speaker wire did not come loose.

AF was able to use the pen to point to various items in the visual scene display, listen to the object's name, and then pronounce it herself (see Figure 6.5). JB asked her to follow this pattern but at times she would continue to touch objects without practicing saying the name herself. Her ability to point to both small (1 cm) and large objects with

Figure 6.5: AF uses the digital pen and interactive worksheets to explore a kitchen visual scene display. Tapping on an object plays the name of that object.

the pen on the visual scene display was excellent. One problem is that she did not know which items in the kitchen scene were active, as many had audio associated with them, but some small items (e.g., a water bottle) did not. AF also had no difficulty using the pen as a writing instrument with the sentence completion and written comprehension activities. She looked at the paper as the pen read a sentence aloud and then filled in the blank or answered the question.

Independent use for homework. JB continued to use the kitchen activity worksheets during another therapy session with AF. She became more proficient at using the pen. She quickly figured out what she needed to do for each activity and began using the pen with less support from JB. At this point, a goal for AF was to determine whether she would be able to use the pen independently in her home for language practice outside of structured therapy sessions. One researcher brought the digital pen and paper kitchen activities to AF in her home to determine how much support she would need away from JB. While AF remembered using the pen with JB, she needed help turning on the pen and getting started with the first activity. For the homework activities, we modified the existing worksheets to include an answer box that would play the correct answer when tapped, as AF was constantly worried about whether she answered a question correctly. AF was able to independently perform a completing the sentences and a written comprehension task with the digital pen; however, she did not use the pen to listen to words that were difficult for her to pronounce or to verify that she answered each question correctly. Even with prompting from the researcher, the concept of tapping on an answer box to hear the correct answer was too complex for AF. We also explored the audio recording facilities of the digital pen. AF read aloud a passage from a novel she was currently reading, and the researcher helped her record her voice using the digital pen. When asked what she thought about the pen recording and playing back her reading, she said, "I'm not doing very well," indicating that she mispronounced many of the words. She was able to understand that the pen could record her voice and help her reflect on her speech. The ability to record and replay speech from a client working independently is compelling, yet the process needs to be extremely simple with explicit instructions.

6.6.2 Case: Augmenting Communication

Three years ago MW (age 90, female) had a stroke in the left side of her brain. The stroke resulted in aphasia and apraxia of speech. She participated in therapy with JB for several months immediately following her stroke, then took two years off, and recently began therapy again with JB. She is able to understand written and auditory information, but she has extremely limited speaking and writing abilities. The stroke made her unable to speak any words consistently or write legibly by hand. People who interact with MW primarily ask her yes/no questions, to which she responds with a one word answer (e.g., yes or no), head nod, or shoulder shrug. JB explained of MW, "Sometimes she can say full words, other times nothing... There are times when you talk to her where she has one or two word utterances. Talking to her yesterday, I asked her a very direct yes/no question and she said 'no.' That's just the nature of apraxia." MW has been in this communicative state for three years, and according to JB, she "is grateful for the help but very frustrated... Her goal was always rehabilitation. 'Get me back my speech.'" Recently, JB decided to shift from rehabilitation to compensatory communication strategies.

Materials. As a communication aid, JB taped a paper picture board with icons and an alphabet board to MW's bedside table. JB has been working with MW to use this as a means of communication with nurses and community staff (e.g., to request medicine). MW can point (with her index finger) to icons on the picture board to indicate that she is hungry or in pain, for example; however, she often points between two icons or just below an icon, making it difficult for the recipient to understand her. At first, she was slow to scan the picture board and locate an item (took up to 45 seconds to locate a single icon), but with a couple weeks of practice her search speed has increased. We integrated the digital pen with her existing laminated picture boards. Now, touching part of the picture board with the digital pen plays the audio associated with that icon or letter. For this use case, the digital pen had a non-marking tip to serve as a pointing device rather than a writing instrument.

Initial use. During the first session of use (approximately 45 minutes), MW was able to easily grasp the digital pen in her hand and use it as a pointing device. While holding the pen her fingers did not occlude the camera under the pen tip nor did she accidentally press the power on/off button. At one point during the first session, MW held the pen at an angle where its camera could not read the location on paper. JB helped her adjust the pen position to a more upright angle. MW understood that touching on an icon or letter played audio associated with that item, as she would look up at JB and nod after hearing the sound. Previously, JB taught her to point to icons with her index finger. As a result MW often pointed with her index finger first then the pen or pointed with the pen and her left index finger at the same time. Using the pen while connected to the external speaker was also not a problem.

An interesting segment of activity occurred when JB had MW practice locating icons. She could not find a certain icon after 20 seconds and began to beat her hand on the table. JB then pointed to the *frustrated* icon and asked, "Are you feeling frustrated?" MW waited for five seconds and pointed with her index finger to the *cold* icon. JB responded, "You are? Press it." MW touched the *cold* icon with the pen. JB then said, "What do you want? Using this tell me what you want." MW touched the *blanket* icon with the pen and her left index finger, then looked up at JB (see Fig. 6.6). JB responded,

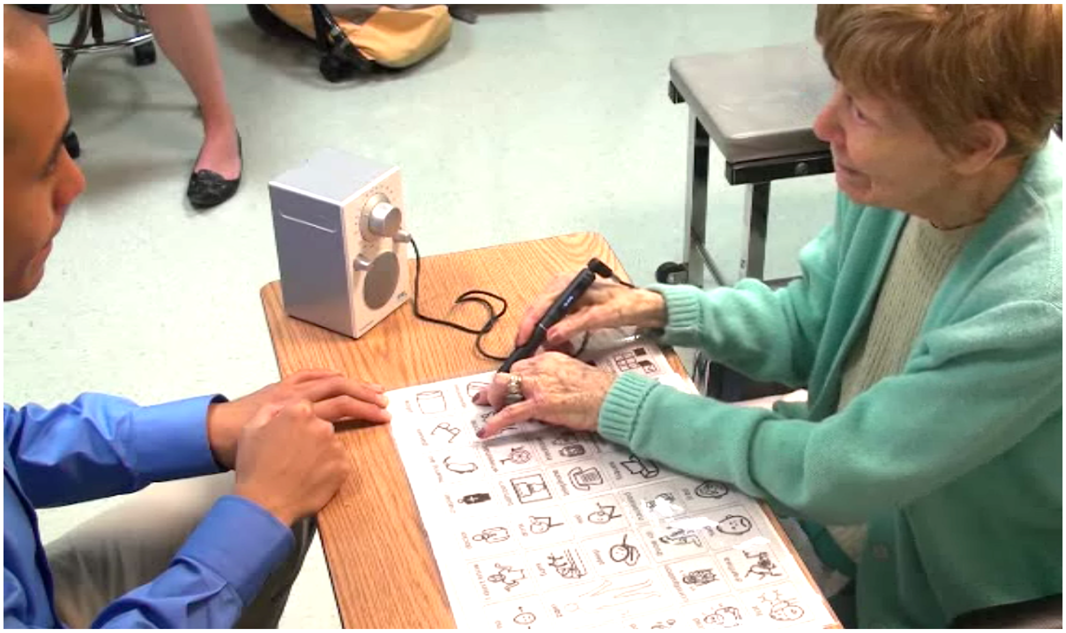


Figure 6.6: MW tells JB that she is cold and would like a blanket. She points to the *cold* and then to the *blanket* icon using the pen and her left index finger.

"OK, I'll go get you a blanket." He said of the experience, "It was great that she was able to tell me what she wanted."

Applied use for social engagement. JB spent two additional therapy sessions helping MW learn to use the digital pen and interactive picture boards. The fourth time MW used the digital pen was during a group activity. JB arranged for MW to call out bingo numbers for fellow residents during game time. Applied use of the communication board in this social situation was challenging for MW. This interaction required her to read the bingo letter and number off the ball and then tap the letter and the number for other players to hear. For the first half of the game, JB provided hand over hand guidance to help MW tap out the bingo letter and number. MW was able to do this on her own by the end of the session. She had difficulty, however, with numbers that required double tapping. For example, the number 44, required her to tap 4, release, then tap 4 again. On two different occasions she held the pen down instead of tapping the number twice. The most difficult aspect of interaction for MW was the visual search required to find a letter or number. JB explained afterward, "I think tracking is an issue for her [MW]. It [the picture board] may be too overwhelming. She would look and then give up." Refining the layout, icon size, and icon images for the alphabet and picture boards may help address this problem. This is an easy modification given the flexibility of the digital paper and pen system, and we elaborate this point in the discussion section below. While the task of using the digital pen to call bingo numbers was challenging for MW, the digital pen helped her to engage socially with her peers. She smiled and interacted with other residents, and they cheered for her when she used the pen to call out a bingo number. MW's stroke and subsequent communication loss has made her withdraw socially, so enabling her to engage in this type social setting is a positive benefit.

Figure 6.7: JB helps MW call bingo numbers using her alphabet board and the digital pen.

6.7 Discussion

The multimodal pen and paper interface is a promising tool for speech-language therapy. Our case studies demonstrate that an 88- and 90-year-old were able to successfully use and understand the digital pen and paper system. Conducting an evaluation at the retirement community allowed us to observe a single therapist over a period of time and how two of his clients responded to the technology as an AAC device. One limitation of this approach is that we were constrained to the therapist's caseload of clients, and at the time of this study he only had two clients who were willing and able to participate in research. Exploring multimodal interactive paper technology for a wider age range of clients is one goal of future work. Based on field observations, feedback from therapists, and case studies of use in an authentic setting, we now discuss advantages and challenges of using digital pens in therapy activities.

6.7.1 Usability Issues

Based on our preliminary case study evaluations, the interactive paper materials are fairly easy to use. JB said, "What you have is excellent. It can be incorporated with residents in it's current state." He continued, "It's relatively user friendly, not so much of a high learning curve. It's just using a pen and pointing, then getting feedback. They get an extra notch of stimulation." Several aspects of interaction require further exploration, such as whether clients would be able to perform gestures with the pen (e.g., double tap). The form factor of the pen worked well for this audience, but we need to explore the prototype with adults who have less motor ability in their dominant hand. The older adults we observed were able to hold the pen successfully without accidentally covering the infrared camera or turning the pen on/off. Occasionally they would hold the pen at an angle almost parallel to the paper, making the pen unable to read it's current location. Future prototypes should have a feedback mechanism (i.e., audio notification) to help users hold the pen at an appropriate angle.

6.7.2 Logistical Challenges

Fitting with the current workplace practices and logistical challenges of a therapist is critical. One challenge with the current device setup is transporting the equipment to different clients. JB explained, "As a therapist, I want to keep it simple. I don't want to lug a bunch of stuff around... plugging things in, going from room to room. I keep going back to a computer. It can all be compact in one." In contrast, one of the clinic therapists said, “I think it would be great for an everyday AAC device because it is so compact.” The amount of equipment a therapist is willing to transport to a client’s location may largely depend on how much the device benefits a particular client and whether or not an external speaker is necessary.

6.7.3 System Flexibility

The system is appropriate for use in structured one-on-one therapy sessions, group therapy, independent language practice at home, and as an alternative communication device an individual could take with them throughout their daily activities. JB said, “I can see her [AF] taking these [worksheets] home with this device to help her. For others, it would be a therapy task... Yeah, it would be great for in home use.” With refinement and additional user training, the interactive paper system would be appropriate for a client like AF to use as a homework tool. MW could carry the pen with her throughout the day with a smaller picture board, and it could serve as an alternative communication device.

Customization of activities allows the digital pen to be useful for a wide range of individuals with aphasia. The flexibility of paper allows for rapid generation of multiple interfaces that can be tried out with clients. This flexibility allows for interface characteristics such as the layout, size, and color scheme to be changed with limited effort. Therapists can also cut out and paste pieces of interactive paper in various ways to rearrange an interface according to a client’s needs.

6.7.4 Device Adoption and Acceptance

As with other assistive devices, adoption depends ultimately on what the client will accept. JB said of MW and her husband, “I showed them DynaVox, and her husband was like, ‘I don’t know if I want a device. I want [her] voice back.”’ JB elaborated, “Acceptance of [a device] is one thing. We must slowly introduce it because this is effectively their voice... It all depends on them ultimately.” The therapists at the university clinic suggested that some of their clients who were resistive to more high-tech AAC devices might find the digital pen a suitable alternative.

6.7.5 System for Content Generation

Currently, a researcher must help the therapist generate pen-based content for weekly therapy activities. This is appropriate for our study because we are just starting to understand the potential of multimodal pen and paper interfaces to support therapy practices. However, given the different needs and requirements of clients, a more flexible approach allowing therapists and potentially family members to define extensible multimodal interfaces is needed. Chapter 7 describes a system for authoring multimodal paper-digital documents based on the early research in this chapter.

6.8 Conclusion

This chapter examined the practices of speech-language therapy and how a multimodal digital pen and paper might support this process. Our analysis presents a first look at the opportunities enabled by this kind of interaction. The two case studies demonstrate the feasibility and appeal of this idea, but the potential for multimodal pen and paper interaction to support therapy processes is seemingly infinite because of the flexible and pervasive nature of paper. Beyond structured therapy activities, an individual could carry a pocket-sized flip book to support daily interaction or print interactive content on stickers or note cards to place in a home or clinic environment. The digital pen's ability to record interaction, both writing and spoken dialogue, makes it a promising device for documenting therapy activities over the duration of treatment. Therapists can automatically log client activity and archive notes about client progress to support assessment, similar to the Abaris system (Kientz et al., 2005). Digital pens can also be combined with computer display systems, and the potential to integrate more dynamic content presented on a multitouch computer with a digital pen is a promising avenue of exploration. The following chapter introduces a system for authoring multimodal paper-digital interfaces as a way of opening up a range of new possibilities and opportunities for therapy.

Chapter 6, in part, is a reprint of the material as it appears in the Proceedings of the 2010 ACM Conference on Computers and Accessibility (ASSETS). Piper, A.M., Weibel, N., and Hollan, J. Introducing Multimodal Paper-Digital Interfaces for Speech-

Language Therapy. The dissertation author was the primary investigator and author of this paper.

Chapter 7

Write-N-Speak: Authoring Multimodal Therapy Materials

This chapter builds on findings from Chapter 6, which examined the process of speech-language therapy for adults with aphasia. This exploration of speech-language therapy resulted in the development of Write-N-Speak, a programmable toolkit designed to help non-technical end-users (e.g., therapists and caregivers) independently create custom interactive paper materials. We evaluate Write-N-Speak through a 12-week clinical field study. The therapist used Write-N-Speak to create custom interactive worksheets, photographs programmed with the client's voice, and interactive stickers on household items to aid object recognition and naming. We conclude with a discussion of multi-modal digital pen technology for this and other therapy activities.

7.1 Introduction

Aphasia is a communication condition characterized by a reduced ability to understand and/or generate speech and language. It often results from a stroke, traumatic brain injury, or neurological condition. Stroke is the third leading cause of death in the United States (National Stroke Association, 2010). The World Health Organization (2007) reports that worldwide, 15 million people suffer stroke each year and Europe averages approximately 650,000 stroke deaths each year. A stroke occurs when an artery of the brain becomes clogged or ruptured and disrupts blood flow to part of the brain.

Brain tissue can die or be injured, causing a change in neurological function. Of the estimated 400,000 strokes which occur each year, 80,000 result in aphasia (The National Aphasia Association, 2011). These individuals may have difficulty understanding written and spoken language and/or be unable to speak fluently or write. A stroke can also result in loss of motor movement in one's face, arm, hand, or leg.

The process of speech-language therapy helps individuals with aphasia regain language comprehension, practice speaking, and find strategies to cope with communication challenges. Speech-language therapists use a variety of techniques and devices to facilitate this process, including many forms of paper-based materials. We describe field observations of speech-language therapy in Chapter 6. As described in the previous chapter, multimodal digital pen technology is promising for supporting speech and language activities. The flexibility, portability, familiarity, and cost of digital pen and paper technology make it an appealing tool for this and potentially other therapy contexts.

In order to further overcome the limitations of paper and exploit the advantages of digital interfaces, the Anoto Digital Pen and Paper technology (Anoto, 2011) enables users to flexibly access digital information through pen-based interactions on paper. Several frameworks have been developed to manage the interactions across the paper-digital divide, supporting manipulations of paper-digital documents and helping developers to build pen-and-paper applications (Yeh et al., 2007; Guimbretière, 2003; Konishi and Ikeda, 2007; Signer, 2008; Norrie et al., 2006). These frameworks allow paper versions of digital documents to be mapped to their original source, thus creating a new kind of enriched paper-digital environment (Weibel, 2009), and enable interactions from pen and paper interfaces to the digital world through gesture-based approaches (Liao et al., 2008; Signer et al., 2007).

One limitation of existing digital pen technology is that it is difficult for non-technical users to generate custom interactive paper materials. This was the case for speech-language therapists involved in this research project. Enabling a broader user group to generate custom paper-digital materials was the main goal of the work within this chapter. Similarly, few computer systems for therapy serve as a toolkit to help therapists create customized therapy materials for individuals with communication challenges. Bungalow Software (Bungalow Software, 2011) is an example commercial product that allows therapists to tailor computer-based speech and language activities to clients' needs, but it is limited to a computer workstation with activities following a predefined script. Another example is SpeechKit, a multimedia system for use by speech therapists to help rehabilitate individuals with motor impaired communication (Calder, 2008). More broadly, there is a need to understand speech-language therapy in order to design tools that better support this process.

In this chapter, we present Write-N-Speak, a paper-digital toolkit for end-user creation of custom therapy materials. We deployed this system for 12 weeks with one therapist-client dyad in a clinical setting. The authoring system is tailored to the needs of speech-language therapists working with adults with aphasia, but the system is built with a flexible architecture that can be adapted to a variety of use cases requiring the creation of custom paper-digital materials.

7.1.1 Design Requirements

We established the following design requirements for a system to support client-therapist interaction based on our understanding of speech-language therapy and initial exploration of digital pen technology. These requirements evolved from field observations in two clinical settings, analysis of video recorded therapy sessions, interviews with therapists, and detailed field notes.

Support face-to-face interaction. Therapy is a collaborative process that involves face-to-face interaction between the therapist and client. Most often the therapist and client sit face-to-face at a desk or table. Eye contact, lip reading, and gestures are critical to establishing a shared understanding of language during a therapy session. The system should support at least two users working closely without occluding multimodal cues that are integral to communication.

Provide multiple, multimodal representations of language. The core skills that therapy addresses are reading, writing, speaking, and auditory comprehension. Therapists use redundancy in language (i.e., a word may be written, spoken, and gestured) because clients vary drastically in how their aphasia impacts communication. A device to support therapy should span these dimensions of language and enable redundancy of language via multimodal input and output.

Leverage existing paper-based practices. While a therapist may try new technologies to help a client, the larger process of speech-language therapy is predominantly paper-based. A wealth of hardcopy resources are available for therapists to use with clients. These include, for example, workbooks with word lists and sentence completion exercises, communication boards, picture books, vocabulary flash cards, and visual scene displays. Systems should support paper-based practices and ideally work in conjunction with this established content base.

Content should be customizable. Therapists spend several hours each week preparing a treatment plan and materials for each client. There is a wide range of communication needs and goals for adults undergoing therapy for aphasia, so a universal solution is not likely to be successful, just as one treatment plan is not appropriate for all clients with aphasia. Instead, a system should allow therapists to create custom content to support individual treatment plans.

Portable and quick to set up. Therapists meet with clients in a variety of locations and often move between these locations throughout a single day. For example, meeting locations include a desk within an outpatient room, a bedside table in a hospice facility or health center, and a kitchen table within a private residence. Furthermore, therapists usually see multiple clients in a day and have limited time to set up for each session. A device that is portable and quick to set up is therefore an important component of the envisioned system.

Given these requirements, our research team considered a variety of devices, including small touch-screens (e.g., Apple iPad), large multitouch computers (e.g., Microsoft Surface), desktop software, and digital pen and paper technology. Among these alternatives, digital pen and paper technology is a promising yet underexplored solution for speech-language therapy. Paper is easy to transport, duplicate, and manipulate, enabling spontaneous collaboration at a shared workspace. The recent introduction of a novel kind of multimodal digital pen, the Livescribe Pulse Smartpen (2011), supports interaction with both written and spoken language. This "pentop computer" captures and recognizes paper-based handwriting while also providing audio recording and playback features. In addition to these multimodal capabilities, Livescribe technology enables custom-made applications and content to be deployed directly onto the pen. Moreover, Livescribe digital pens are relatively inexpensive (starting from $99).

7.2 System Design

Existing digital pen and paper technology (e.g., Anoto, 2011) enables static paper documents to be turned into interactive interfaces that can be used to recognize handwritten information, link to digital media such as sound or voice recordings, or interact with applications on a separate computer. This technology enables users to exploit rich digital services while keeping the natural interaction common in traditional pen and paper interfaces. However, the creation of applications and corresponding paper documents for digital pens requires programming experience. Our lab is developing a flexible toolkit to support a broader range of end-users in authoring custom multimodal documents. The Write-N-Speak system described in this paper is one instance of the larger toolkit.

Write-N-Speak runs on a Livescribe Pulse digital pen. The Livescribe digital pen integrates a processor, memory, and an infrared camera able to interpret a unique dot-pattern printed on standard paper. By decoding the printed pattern, the device can track the pen's position on paper in real-time. This digital pen is a commercial product with a built-in speaker, microphone, OLED display, audio jack, and USB connector. The Livescribe digital pen and paper system is unique in that it allows development and deployment of standalone applications. Livescribe provides a software development toolkit that is an Eclipse plug-in requiring development in Java Micro Edition (J2ME), and Write-N-Speak is built using this toolkit. Write-N-Speak works with any *open paper* document—interactive paper printed with the Livescribe dot pattern that is detected by any Livescribe pen—including all Livescribe notebooks, notepads, and journals sold in stores.

Write-N-Speak allows therapists to create multimodal paper-digital documents that address the four core components of speech-language therapy: reading, writing, speaking, and auditory comprehension. Therapy activities often involve articulating names of items within a visual scene, written and oral completion of sentences, speaking pairs of words that sound alike, listening to the name of an object spoken aloud and writing its name, and recording one's own speech for self-reflection. All of these activities are supported by Write-N-Speak. There are three aspects of authoring documents with Write-N-Speak: (1) configuring paper materials, (2) programming multimodal interactions, and (3) using the custom content.

7.2.1 Configuring Paper Materials

A key feature of Write-N-Speak is that it allows users to write on, cut, glue, print, reprint, and manipulate a variety of paper materials, thus preserving the affordances of paper that make it such a wonderful medium for a range of activities. Many laser printers are capable of printing the high-resolution dot pattern with specialized ink required by the Livescribe digital pen. Users may write directly on paper printed with the dot pattern, print content such as photos or images with a dot pattern overlay, or if they do not have a high-resolution printer, they can print over a page of pre-printed dot pattern paper with any standard desktop printer.

Write-N-Speak also allows users to print the dot pattern on opaque and transparent stickers, and we experimented with Avery (Avery, 2011) brand stickers available at any office supply store. Stickers enable multimodal interaction to extend to a variety of resources, including paper artifacts without the dot pattern (e.g., books and magazines) and physical objects. Individual pieces of paper and stickers may be cut up with scissors and reconfigured with tape or glue. For the case study below, the research team provided the therapist with a set of paper materials, including blank pages of paper, notebooks, photographs, and various size stickers all printed with dot pattern.

7.2.2 Programming Multimodal Interactions

The main contribution of Write-N-Speak is its pen-based interface for creating multimodal interactive paper materials. Write-N-Speak has two modes of use: *paper editor mode* for authoring multimodal content, and *client mode* where individuals with aphasia may use the newly created materials. To add content to interactive paper, a user starts the *paper editor* mode on the control panel (see Fig. 7.1). The software provides users with audio instructions that walk them through the steps of adding multimodal content. This involves drawing a new region or selecting an existing one on dot pattern paper, defining the multimodal content to be assigned to it (e.g., a button that plays a specific recorded voice), and saving the region.

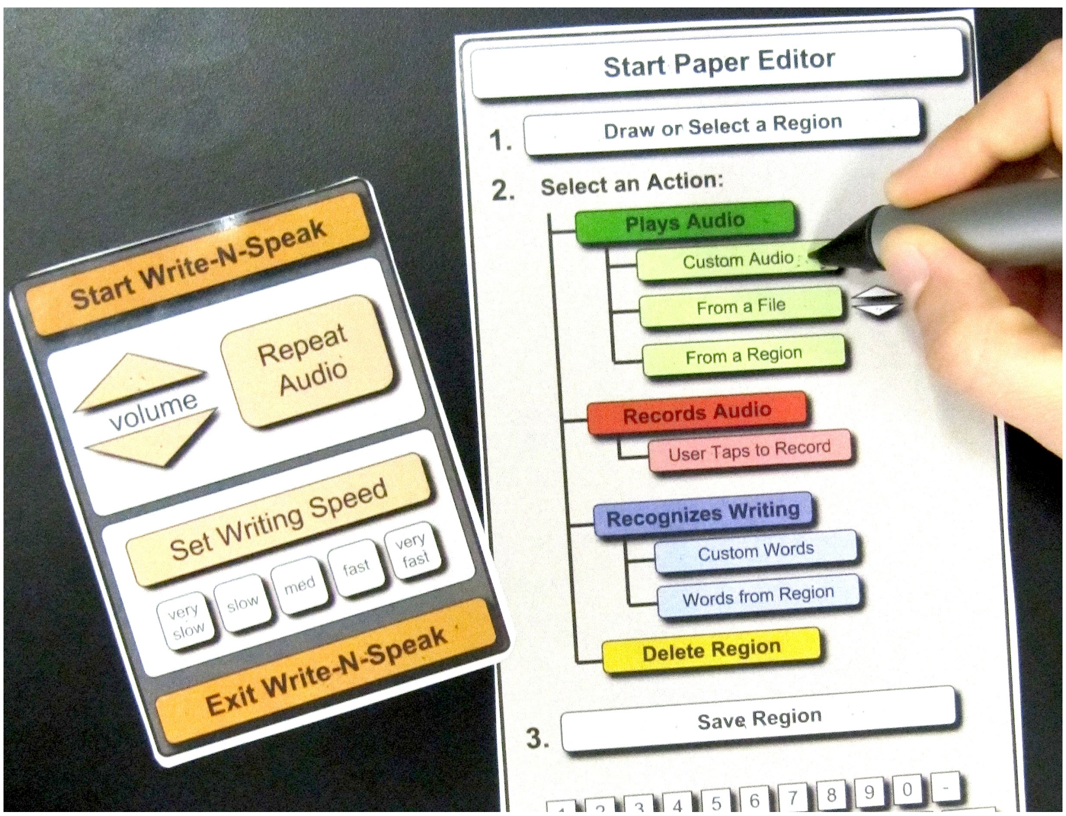


Figure 7.1: Main control panel (left) for use by client and therapist during a therapy session. Paper editor control panel (right) enables the therapist to program the digital pen with custom multimodal content.

Draw or select a region. First, a user defines a region for interaction and multimodal content. They do this by drawing a shape on any paper with the Livescribe dot pattern or double tapping on an existing region that has previously been defined and stored on the Livescribe pen. Stickers already contain pre-established regions, so the user just double taps on them to activate or link specific actions to the sticker.

Define the pen's action. Three types of multimodal content are supported by Write-N-Speak. These include audio playback, audio recording, and handwriting recognition. First, users can designate regions to play audio. Audio may be custom recorded by the user via the pen's integrated microphone or selected from a list of preloaded system files. Additionally, an audio region can be linked to another Write-N-Speak region containing a specific audio file so that the new region will play the previously recorded audio file. This last feature is particularly useful for creating *play* buttons that play audio files associated with other interactive regions. Second, a user can add a region that acts as a *record* button. By combining record regions with play regions, users can create record-and-play buttons. In either case, the recorded audio is saved onto the pen for future use. Third, users can program a region to recognize certain letters or keywords when written by hand with the digital pen. Write-N-Speak exploits the intelligent character recognition (ICR) software deployed on the Livescribe pen and a pre-defined lexicon to detect handwriting. The control panel provides users with a keyboard where they can add words to the pen's lexicon in real time.

Save the region. Once the region has been created, the user must save the region via the control panel. Regions and their corresponding actions can be redefined or deleted any time. As the user defines regions, they become layered on the interactive paper interface with newly added regions occluding existing regions. This allows users to create nested interactions and define subregions within regions. The user must exit the *paper editor mode* to interact with the new multimodal content. Write-N-Speak saves and reloads all user-added content between each session of use. The pen provides step-by-step audio instructions on how to add multimodal content as well as other common tasks, such as drawing a region and using stickers. This option can be disabled by experienced users.

Write-N-Speak also allows users to specify pen-based interactions (e.g., single or double tap) for specific regions by either demonstrating or selecting the user action. However, feedback from therapists and observations of clients with aphasia indicate that double tapping (and sometimes single tapping) is too difficult for certain clients. For this reason, in the version we deployed during our evaluation, all actions are executed on the *pen-down* interaction by default, but the paper-based control panel can be easily modified to allow customization of user interaction and even define multiple interactions per region (e.g., single tapping a region may record audio while double tapping the same region may play the saved audio file).

7.2.3 Using Custom Content

After the therapist has configured the paper documents and added multimodal content, the materials are ready for use by a client. The therapist and clients use the main control panel (Figure 7.1, left) that allows them to start and exit Write-N-Speak, adjust the pen's volume, repeat the last audio, and change the speed at which handwriting is recognized by automatically tuning the waiting time before ICR is executed (this feature improves handwriting recognition and is important because individuals with aphasia have varying levels of writing ability).

7.3 Field Deployment

Through a field deployment of Write-N-Speak, we aimed to: (1) understand the learning curve of Write-N-Speak for an adult with aphasia and their therapist, (2) determine how a therapist might integrate the system into the ongoing treatment of a client with aphasia, and (3) identify usability issues and additional features for future use.

7.3.1 Study Participants

We evaluated Write-N-Speak with a single therapist (JB) and client (AF) over a 12-week period. For more than six months leading up to this field deployment, our research team conducted weekly observations of JB interacting with his clients, each week conducting one to three hours of observation. JB has been a licensed and practicing speech-language therapist for four years. He works at a retirement community and at other local senior care facilities. In this setting, we observed JB conduct therapy with several older adult clients, but the field deployment of Write-N-Speak focuses on JB's interaction with one client (AF, age 88) during weekly hour-long therapy sessions.

AF had a stroke in the left side of her brain approximately one year before we began field observations. As a result, she has aphasia and apraxia of speech that make it challenging for her to construct complete or coherent sentences and articulate words. However, AF is fairly high functioning – she is able to say many words and generally able to communicate her needs. She is an avid reader, and her stroke only had minimal impact on her ability to process written information. AF lives alone, but is well known throughout the retirement community. She often stops to greet other residents and asks "How are you, dear?" Her communication challenges become more evident during structured tasks where she needs to follow instructions or process competing auditory stimuli. She has fully recovered her motor ability in her hands although she uses a walker for balance.

We chose to focus on this therapist-client dyad for several reasons. JB has a strong interest in assistive devices and was willing to incorporate a first version prototype into therapy. AF often states that she will try anything that might help her or help others. Before her stroke, she regularly used a computer, cooked, and made jewelry. This dyad worked together for therapy for one year prior to our research team starting this project, and this established relationship allowed JB to explore new ideas for therapy. Observations of JB and AF at the retirement community enabled our research team to understand therapy practices over time and the progression of therapy techniques as AF's communication needs evolved.

JB was instructed to use the digital pen and materials however he chose with his client. We told him the study would last several months and he could decide when and how to introduce the system into his therapy plans. We provided JB with the following materials: one Livescribe Pulse digital pen loaded with Write-N-Speak software, 20 blank sheets of paper with dot pattern, 4 pages of stickers with dot pattern (white and transparent, varying sizes), interchangeable ink and non-marking pen tips, a pen charger, and an external speaker to amplify the pen's volume if necessary. Before the 12-week study began, JB, AF, and one researcher took photos of meaningful items within AF's apartment. AF pointed out pieces of artwork and keepsakes while the researcher took photos of these objects as well as photos of her living space. The researcher then printed these digital images on paper with the dot pattern and included them with the material set.

One researcher attended and video recorded four therapy sessions over the 12 week study. We wanted the therapist and client to independently decide how to use Write-N-Speak without researcher oversight. Segments of video data were later transcribed and analyzed to understand usage patterns and usability issues. The Write-N-Speak application logged all interactions with the pen, and this data provided insight into therapy sessions the research team did not observe. Additionally, JB described his use of Write-N-Speak in a journal.

7.3.2 Usage Patterns

A valuable feature of digital pen technology is its ability to log user behavior, including handwritten content, recorded audio files, and general interaction with paper-based interfaces. This data could help a therapist understand what a client is doing independently at home, as well as document a client's therapy progress over time. For the purposes of this case study, the Write-N-Speak application logged pen usage data to help understand how often the therapist and client performed certain actions. Over the 12-week study, JB used Write-N-Speak for 349 minutes (123.4 minutes preparing therapy materials and 225.6 minutes during therapy sessions with AF) and six times during therapy. Figure 7.2 indicates weekly usage patterns.

Interestingly, JB almost always created custom audio playback regions and rarely created record regions or handwriting recognition regions. Over the 12-weeks JB and AF created 94 regions with custom audio and accessed them a total of 423 times. Record and handwriting recognition regions were created and accessed less than five times each over the study period. Instead of creating record buttons and having AF use them in *client* mode, JB helped AF create custom audio regions with the *paper editor* mode so

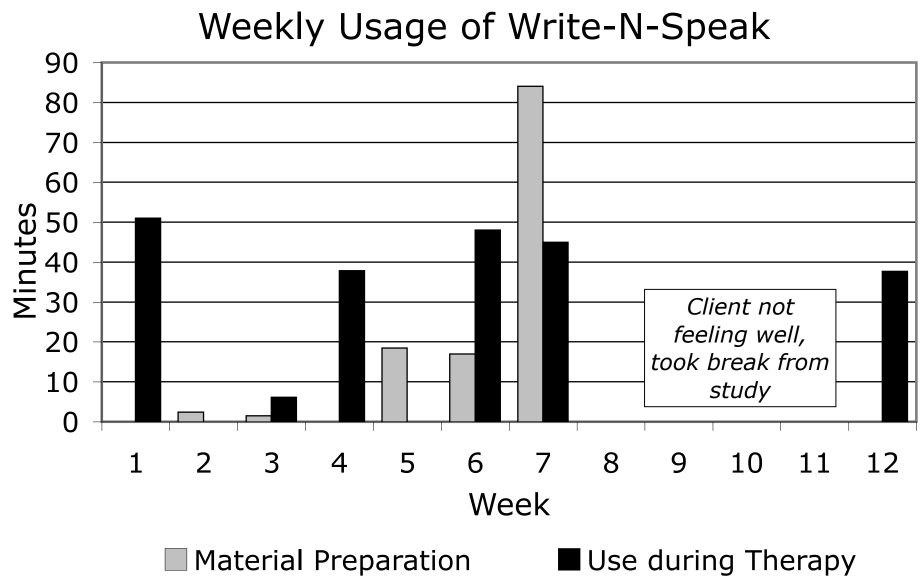


Figure 7.2: Time spent preparing materials and using Write-N-Speak. Total usage time is 349 minutes (123.4 minutes preparing therapy materials and 225.6 minutes during therapy sessions with AF).

she would hear her voice by tapping on various parts of the paper interfaces. This finding indicates a need for a simplified control panel that would allow JB and AF to quickly create play and record regions.

7.3.3 Learning to Use Write-N-Speak

In Week 1 of the case study, one researcher introduced Write-N-Speak to JB during a routine therapy session with AF. AF and the researcher observed as JB learned to use Write-N-Speak. This session was video recorded and later analyzed to document usability issues and JB's feedback on the system. JB figured out how to use Write-N-Speak to add play and record buttons, link a play and record button, designate an area to recognize handwriting, and delete a region. Throughout this session of use, he explained what he was doing to AF and the researcher. He commented, "I was initially saying, 'Wow this is a lot of steps.' But then I was saying to myself, 'Okay, as I'm going through the motion of it, it gets a lot easier."' The order of the different steps confused JB at the beginning and he had difficulty remembering to save the regions he had created.

Noting both the creativity and customizability that the Write-N-Speak application allows, he said, "You just have to learn and then you can figure out creative ways to implement it." JB also commented on the main advantage and design feature of Write-N-Speak: "You can customize it [paper materials] to make it really tailored to whatever you feel is appropriate for the resident." Write-N-Speak is designed to help therapists like JB create custom activities that suit the idiosyncratic needs of clients. "A lot of times, you have an activity that's limited to one kind of person with a certain disorder, but now I can manipulate it in the sense to make it function like this," he elaborated. With Write-N-Speak, JB explained, he can make activities suitable for the specific needs of any client. JB repeatedly stated that an important benefit of the Write-N-Speak application over traditional pen and paper is that that it offers "another dimension" for therapy activities. He commented, "I just like how you can make a picture more dynamic [by adding] voice to it... I think it's a great tool to add more cueing to just regular worksheets." JB continued to explore Write-N-Speak on his own during Week 2 before introducing it into therapy.

7.3.4 Adding Speech to Client's Photos

During Weeks 3, 4, and 6, JB used Write-N-Speak to add audio content to a photo of AF's living room. His goal in using Write-N-Speak was to "develop word-finding tasks to label nouns." It took JB and AF three therapy sessions to develop a successful process for labeling items on the interactive paper. In Week 3, JB added his own voice to various items within her living room picture (e.g., couch, television, telephone) and had AF use the pen to listen to the names of objects. AF said this activity was "too easy," so the following week JB decided to have AF record her own voice for objects within the living room picture. This was challenging for both AF and JB, and he said of the experience, "It was difficult for AF to understand what was expected of her during the session." JB decided to take a break from Write-N-Speak for Week 5. Although JB quickly figured out how to author multimodal therapy materials, it took him several weeks to figure out how to design activities that were at an appropriate level for AF.

In Week 6, JB successfully helped AF program the pen with her voice. To do this, JB started and stopped the recording on the paper editor control panel, pointed to the name of an object written on paper, and then held the pen up toward AF while she said the name of the object. The following segment and Figure 7.3 illustrates this process:

JB: *Alright, [Mrs. F]... I want to program this pen to have your voice on it, ok?*

AF: *Ok.*

JB: *...What I'm going to do, and you remember, I did this last week. I'm gonna make, um, squares and I'm going to hold the pen up to you and I want you to tell me what this is.*

AF: *Oh, I see.*

JB: *(points to a chair) This is what? What is this?*

AF: *Wooden chair. A wooden chair.*

JB: *Ok, good.*

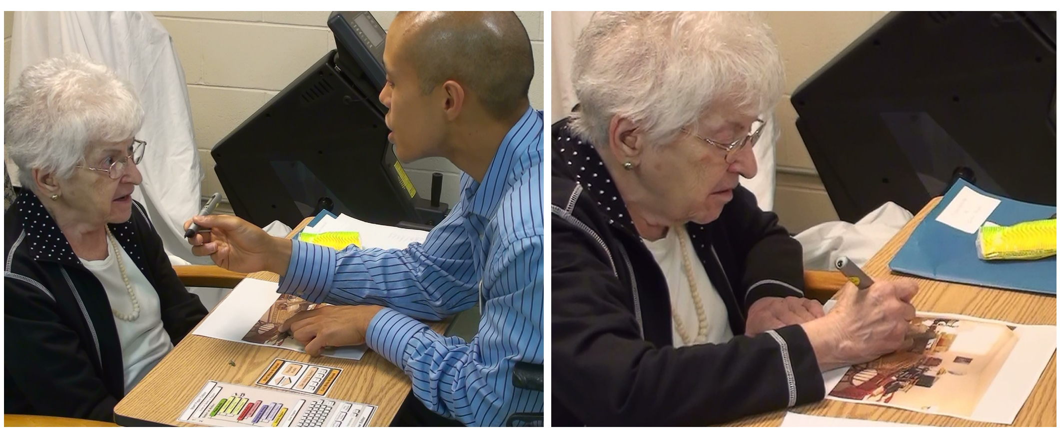


Figure 7.3: *Left*: JB helps AF add her voice recordings to a photo of her living room. JB helps her pronounce the object name if she has difficulty. *Right*: after adding audio for all objects in the living room picture, AF uses Write-N-Speak to listen to how she pronounced the name of each object.

JB said of this experience, “I’m so happy, finally we got a system going. She actually is doing it correctly... which is good... The activity that we’re making is going to be helpful. It will be helpful for her, but I think just the process right now that we’re doing to get her to follow directions... I think that in itself is helping her.” JB explained, “It’s good to see that over the course of the sessions, three or four sessions we’ve been working... we were doing exactly this, trying to have her program it herself, and that’s what I want. And I wanted her to be part of it, because I wanted her to get a feel of what we’re trying to accomplish here.” About the earlier sessions, JB said, “I know she was frustrated and I was getting a little frustrated towards the end... I needed to really refine the task. So then the next time around, I touched base on it, but then I said, you know what, I’m going to program it,” so JB programmed the living room picture with his voice. He started the therapy session during Week 6 with his recorded audio but then reassigned the existing regions to incorporate AF’s voice.

JB said of AF, “...she knows that when it’s successful, it’s successful, and she’ll make these comments. And she knows when it’s not, she’ll start saying, ‘oh this is stupid,’ or she’ll say, ‘this is too easy.”’ Reflecting on Week 6’s session, JB said to AF, “Today was a really great...lesson because it showed me that, not just from you, but that people can learn. That they can really learn how to overcome these little obstacles by just, repetition, just keep repeating. Because remember the first two times we were doing this? We were getting frustrated, right? Remember?” AF responded, “Well, well that’s because you’re, you’re helping me.” JB continued, “So, today, look, we were able to get the whole list done. We were able to... label all of them.”

After AF recorded the object names, JB instructed her to say the name of each object aloud and then touch the pen to the object to listen to how she said it earlier (see Fig. 7.3, right). The audio recording and playback facilities enable this type of reflection, a central aspect of therapy. Later AF commented, “And this is for people, for me, to, to, to understand what I should sound like.” With respect to audio recording, JB said, “It just gives...personalization to the task. It’s her voice. She’s listening to it and she’s doing the task...” He elaborated, “If you’re part of the process, you know, you will understand. You’ll find it more rewarding because you helped make it... You’re empowering her.”

7.3.5 Interactive Stickers on Physical Objects

For Week 7, JB prepared a different therapy activity with Write-N-Speak. Again, the focus of this activity was word finding and object naming. JB added transparent stickers (printed with the Livescribe dot pattern) to a variety of household objects. He programmed each sticker to say the object name when tapped with the pen.

During the therapy session, JB asked AF to recall the name of each object and then use the digital pen to listen to the object name read aloud. This was a difficult task for AF to understand, and JB explained the activity to AF multiple times. At one point, AF held the pen up to her mouth as if to record speech, and JB said, "I think you're mixing up what we did last week." Later, AF tried to write with the pen on the sticker instead of just tapping it. JB and AF worked on the object recognition task for 40 minutes and then decided to stop. He said, "[AF] is too confused. This is all learning for us." JB asked AF if using the pen was too hard for her, and AF admitted that she was frustrated. JB responded, "This is part of speech therapy. This is why you are coming to speech therapy." She then said, "I guess I'll try it again." JB continued, "You're having challenges and that's part of the whole process. I don't think you're frustrated from the work..." and suggested that the challenges of her stroke were the frustrating part. "We just have to keep going," JB said, "Every day you're getting better. Every week you're getting better." They decided to take a break from Write-N-Speak for next week's session. Afterward, JB explained that it's important to keep pushing AF with new activities but not to let her get too discouraged. In hindsight, JB realized that around this point in the study AF had several smaller strokes. The strokes were minor enough to not be noticed immediately by her family or caregivers. JB did notice her change in behavior during therapy but until several weeks later he did not know what caused the behavior change.

In terms of his own learning, in Week 7 JB described programming the pen as "really easy" and stated, "programming is pretty simple now." Write-N-Speak and the transparent stickers allowed JB to explore a novel use of digital pen technology in that he was able to design therapeutic interaction with objects of varying shapes and materials (wood, metal, and plastic). JB encountered two problems with the stickers that are limitations of the digital pen technology. First, the Livescribe pen cannot recognize the

Figure 7.4: *Left*: JB added transparent stickers to physical objects (e.g., flashlight, spoon, cup) and programmed Write-N-Speak to say the name of each object when tapped. *Right*: AF tries to say the name of the object on her own, then she listens to it by tapping the transparent sticker.

dot pattern printed on a black background, and this includes placing a transparent sticker on a black object. JB reported that he could not get the pen to recognize a sticker when placed on a pair of black sunglasses. Second, the pen had difficulty recognizing a sticker wrapped around a small wooden pencil due to the curvature of the object.

7.3.6 Integrating Existing Therapy Content

The design of Write-N-Speak supports therapists in transforming existing paper or digital documents into a multimodal experience for the client. In Week 8, JB used a standard computer printer to print the Rainbow Passage (Fairbanks, 1960) on a blank sheet of paper with the Livescribe dot pattern. The Rainbow Passage is commonly used by speech-language therapists to test an individual's ability to produce connected speech. Sadly, AF experienced several minor strokes during Weeks 9-11 that left her feeling poorly, so JB decided to postpone use of Write-N-Speak until she felt better. In Week 12, JB returned to using Write-N-Speak with AF and incorporated the diagnostic materials he prepared earlier. JB sat across from AF and helped her record each sentence of the Rainbow Passage onto the pen. After recording all of the sentences, JB helped AF review how she read each sentence. The goal for this exercise was to have AF reflect on her pronunciation and articulation of these phrases.

With JB's guidance, AF successfully recorded her voice for ten sentences. After the session JB commented, "She actually did well, she picked it up quickly... It was very successful for her, and I think she felt pretty good about herself." The process JB and AF followed for this week's activity was similar to that of Week 6, and it was positive to see AF remember how to use the system even with a four week break. AF understood that her voice was recorded on the pen. As she listened to her speech for each sentence, JB asked her what she thought of her pronunciation, and AF was able to determine when she pronounced a word within a sentence incorrectly. For example, after AF and JB listened to how she read a sentence, AF said, "I think the other was better. The second was, I was, was, token." JB responded, "Ahh haa, you noticed that, right? You didn't say 'token.' Let's read it again." A nice affordance of digital paper is that the therapist can file away this sheet of paper with other paper documentation. The therapist could then use the digital pen to access the client's speech at that point in time and review her overall progress.

7.4 Discussion

Write-N-Speak allows therapists to author custom multimodal paper-digital documents and explore the possibilities of digital pen technology for therapy. Through the 12-week field deployment, we found that the the therapist and client were able to learn to use Write-N-Speak and that the therapist found creative ways to exploit multimodal pen technology in his therapy plans. Unfortunately, the client's condition worsened during the study and this caused us to stop observations before other aspects of the system, such as handwriting recognition, were evaluated. Nonetheless, this first evaluation of Write-N-Speak along with prototype reviews from additional therapists at the university clinic provide insights for improving the system before we evaluate it with additional populations.

In designing Write-N-Speak, we tried to balance toolkit openness with efficient techniques for programming core therapy tasks. The main challenge of Write-N-Speak, according to JB, is the time required to learn to use the toolkit and prepare materials. Reflecting on the 12 weeks of use, JB said, "The challenging part is just the programming, and I think that's just the learning curve for the clinician and [AF]. I think it's something that has to be clinician guided, very closely, there are too many...little mini steps that you need to do to get the result." While JB learned to create materials with Write-N-Speak, this is not something that AF could do on her own. Beyond learning to use Write-N-Speak, another concern is the time required to prepare materials. JB commented on this, "It takes a little prep time, but sometimes as a clinician you need stuff that's quick, that you can just pull out and work with. Today [Week 12] is the first time where I didn't prep and it was on a whim... and it was pretty successful in that we were able to get a good therapy session out of it by having her read, record, and follow directions." The content creation process could be improved by including shortcuts or paper templates for frequently performed actions. For example, JB helped AF add audio to worksheets by using the *paper editor* interface and rarely used record regions, a feature designed specifically to support clients in recording speech. Creating shortcuts such as pre-linked record and play button stickers would enable a client to more efficiently record their voice and associate this audio with a particular part of a paper interface.

One concern about pen-based technology for speech-language therapy is the quality of audio recording and playback facilities. High-quality audio recording and sufficiently loud, clear audio playback are critical features for digital pens used as an alternate communication device. In the case study, we observed that JB had difficulty figuring out how to best record audio. He experimented with holding the pen's microphone at various positions and found that holding it close to his mouth when recording would provide the clearest and loudest audio for playback. Ensuring that the device has clear audio playback is key for supporting speech-language therapy, as practicing reflection on one's own speech and listening to correct pronunciation of words are core therapy tasks. The pen's volume is another concern, and we provided JB with an external speaker to amplify audio if necessary. He held the pen close to the client's ear instead of using the external speaker during the case study evaluation because the speaker setup was "too cumbersome" for his client. He wanted a wireless audio solution if the pen's integrated speaker could not be improved.

Adults with aphasia vary in their ability to write and manipulate a pen-like device. In the case study, AF was able to use the digital pen to write and point to parts of a paper interface, but she was unable to perform more complex gestures, such as double tapping. Nonetheless, JB, as well as therapists from the university clinic, want future versions of Write-N-Speak to incorporate layered information that is accessed by various gestures. JB explained, "Is there any way to program...a word to do several functions? Like if you swipe across, it will say the word. If you tap it, it will break it down into syllables... and if you even press just the letter, it will give you the phoneme." Our software supports this level of interaction but the current control panel does not provide users with this option. We can easily edit the paper control panel interface so therapists are able to assign multimodal content to different gestures such as pen down, single tap, double tap, and a drag or line gesture. While including this functionality is a minor adjustment to the control panel, it may be challenging for therapists to design activities with gestures that are appropriate for each client's abilities.

7.5 Conclusion and Future Work

This chapter presented the design and initial evaluation of Write-N-Speak, a pen-based interface that allows end-users to author multimodal paper-digital materials. This work is part of a year-long field study that examines the practice of speech-language therapy for adults with aphasia. Through the 12-week case study described in this chapter, we observed novel uses of pen-based technology for speech and language activities, such as recording and reflecting on a client's speaking abilities, as well as adding transparent stickers to physical objects to support object name recall. The study allowed us to begin to understand the utility of digital pens for therapy, but the effectiveness of pen-based technology for treating adult aphasia is still unknown. One goal of our future research is to evaluate how Write-N-Speak impacts specific therapy goals (e.g., improved speech or reading comprehension). We plan to deploy Write-N-Speak with a larger group of therapist-client dyads to examine its effect on therapy outcomes over time. We are partnering with therapists at the university-based clinical training program to enable this level of evaluation.

With improvements to subsequent versions of the toolkit, Write-N-Speak stands to help a wide variety of users create custom activities that leverage digital pen and paper technology. JB said, "I still believe this would be helpful for a lot of people... It's good to see someone like [AF] who's at this end of the spectrum both communicationwise and perception-wise and have a good idea of how the pen can work even with the most severe aphasia cases.” JB also works with younger clients with aphasia as well as at a school for adolescents with autism. He plans to incorporate Write-N-Speak in his work with these two user groups. Although we concluded the case study at 12 weeks, JB requested to keep the digital pen and materials to continue using the system with other clients. Other user groups also stand to benefit from this approach. JB said of the system, “It’s a tool that’s so versatile that you can adjust it to any level... [from] aphasia, cognitive communication, mild onset of dementia, to teenagers [with autism].” Custom multimodal paper materials also stand to impact language learning, both English and foreign languages, as well as improving literacy. One early childhood education teacher described Write-N-Speak as “not only an excellent teaching tool, but a great way for students to reinforce instruction independently.” She continued, “This software would be particularly useful to my students who are learning the basics of literacy and math and often need an adult to explain to them what a word, letter, or even number is. Using the pen would allow students to interact with text and pictures on a page in a way that still actively engages them, yet does not require constant teacher supervision.” The paper-based approach and multimodal features make Write-N-Speak a promising tool for supporting a variety of language learning and development activities.

Our ongoing work involves development of a larger toolkit to help non-technical end-users generate their own multimodal paper materials. Write-N-Speak is only one instance of what can be built with this toolkit. We are refining user interface elements and exploring this authoring toolkit with a broader population of users. Prior to a widespread deployment, a necessary next step is to support sharing of custom content between therapists, or between digital pens. Currently, a researcher must help therapists transfer content between digital pens. We are building a web-based system to aid creation, sharing, and management of multimodal paper-digital documents. The web system will include software to help therapists customize and print control panels for specific tasks. Ideally, the web system will support various communities of practitioners who create and share interactive paper materials.

In summary, Write-N-Speak demonstrates how current digital pen technology can support speech-language therapy. We highlight a range of related communication contexts that may also benefit from this approach. By enabling non-technical users to create and manage their own paper-digital materials, we are opening up digital pen technology to new user groups and supporting flexible content creation by domain experts.

Chapter 7, in part, has been submitted for publication of the material as it may appear in the ACM Transactions on Accessible Computing (TACCESS). Piper, A.M., Weibel, N., and Hollan, J. Write-N-Speak: Introducing Multimodal Paper-Digital Interfaces and an Authoring Toolkit for Speech-Language Therapy. The dissertation author was the primary investigator and author of this paper.

Chapter 8

Conclusion

This thesis presents research on novel systems that support collocated interaction for populations with communication challenges related to a hearing, speech, or developmental disability. The goal of this body of research is to understand how we should design computer systems to augment collocated interaction involving individuals with varying communication abilities, and in turn, understand how these new systems influence collocated communication. In this chapter, I briefly review the three projects presented in this thesis and then discuss the broader contributions of the work.

8.1 Summary of Thesis

The design case study of SIDES (Chapter 3) examines how interactive tabletops might support positive social experiences for adolescents with Asperger's Syndrome (AS), an Autism Spectrum Disorder. Children with AS often have difficulty understanding accepted social conventions, reading facial expressions, and interpreting body language. Through a field study involving a middle school social skills therapy class, we found that cooperative games are motivating for this user population and grounded our system design in this type of activity. SIDES works by exploiting the capacitive input of the DiamondTouch table, requiring the group to make decisions about the state of the game. By having the system control certain aspects of game play, such as turn-taking, the therapist may attend to higher level group issues instead of policing behavior. Several projects in HCI have since built upon this early work using tabletop technology to facilitate social learning and group work experiences for children.

Research on SSI (Chapters 4 and 5) examines the critical interaction space of doctor-patient communication with an emphasis on supporting Deaf patients. A central challenge for many Deaf patients is that they must rely on a sign language interpreter to facilitate communication, sacrificing privacy and independence. Similarly, for doctors interacting with Deaf patients, the presence of an interpreter limits doctor-patient eye contact, an important channel for monitoring the patient's understanding and emotional state. The design of SSI enables face-to-face communication between a doctor and Deaf patient while providing an interactive transcript of conversation on a shared horizontal display. An in-depth analysis of doctor-patient interaction through SSI indicates that the system allows the dyad to make eye contact and allows the doctor to use gesture in communication without competing with the interpreter for the patient's visual attention. Furthermore, SSI transforms the ephemeral nature of face-to-face dialogue into a persistent and manipulatable form.

The Write-N-Speak system (Chapters 6 and 7) is designed based on a contextual understanding of the practice of speech-language therapy involving older adults with aphasia. Speech-language therapy is a cooperative process that involves complex, multimodal constructions of language. Through a year-long field study, I analyzed this process and introduced digital pen technology as a new way of supporting interaction. At the outset of this project, I anticipated creating a multimodal tabletop application to support interaction; however, field research revealed an established workflow involving paper documents as well as familiarity with pen and paper interaction by older adults with aphasia. Through this contextual design study, I innovated new techniques for supporting face-to-face speech and language therapy activities with a hybrid paper-digital system.

8.2 Broader Contributions

Research in this thesis contributes to cognitive science and HCI in several ways. First, I characterized the challenges various populations experience in face-to-face communication, such as limited privacy, independence, and effectiveness in social interactions. I examined pairs and small groups of people interacting in a collocated setting, particularly where one person or more had a disability that impacted communication. Through theories of distributed cognition, I examined the larger system of interaction, including multiple people, their individual and collective tasks, digital and material artifacts, the arrangement of the workspace, and the cultural practices various people bring to the experience. The needs and constraints of each person involved in interaction were considered. This yielded a complex tradeoff space that had to be understood before designing an augmentative communication system. In the SIDES project, children with Asperger's Syndrome needed a motivating experience that encouraged them to engage in positive social interaction. The social skills therapist needed help policing behaviors so he could attend to higher level group work issues. With SSI, Deaf patients wanted privacy and independence in communication. The doctor highlighted the challenges of interaction, primarily attention through eye gaze, with Deaf patients when an interpreter was present. Finally, with Write-N-Speak, older adults with aphasia needed a familiar and easy-to-use communication aid. Speech-language therapists wanted a solution that allowed them to create custom activity materials, the vast majority of which are paper-based, but also required a tool to support direct one-on-one therapy practices.

For each project, I introduced a novel system that supported face-to-face interaction for the population and context of study. System design considered current work practices, tradeoffs between various stakeholders, and how these systems will be used in actual workplace settings. These systems support natural forms of interaction through gesture and/or speech input on shared table and desktop workspaces. SIDES exploited the capacitive coupling of a multiuser tabletop display to influence group work behaviors. As part of this work, I introduced novel user interface elements, such as voting buttons and turn-taking mechanisms, embedded in the cooperative computing workspace. SSI introduced the idea of reifying speech on a shared multitouch display. This feature allowed collocated participants to review, manipulate, and reflect on prior conversation. Reifying speech on a shared workspace has important consequences for all conversation participants, not just the Deaf user. In the Write-N-Speak project, I designed and implemented a pen-based toolkit to help non-technical end-users easily generate custom paper-digital materials. The recording and playback features of digital pen technology, made accessible by the Write-N-Speak system, allowed therapists to record, replay, archive, and repurpose speech. There are many extensions of the ideas presented in this thesis that my future work will examine.

Lastly, my thesis work understands the ways in which novel collocated communication systems shape human interaction. By leveraging theories of distributed cognition, I evaluated human interaction as a system, which includes attending to the role of the body, material workspace, and the social and cultural factors that are important to various populations. I employed a range of methods to establish a holistic understanding of interaction, including laboratory and field evaluations, micro-analysis of video data, in-depth contextual interviews, and behavioral observations of system use. For each system, the goal was to support effective face-to-face communication; however, "effectiveness" in communication was defined slightly differently for each system given the needs of the user group. With SIDES, a positive outcome was sustained engagement by adolescents and positive group actions, such as encouraging other group members. Effectiveness in communication between the doctor and Deaf patients with SSI was characterized by increased eye gaze between the dyad, the ability for the doctor to express herself multimodally, and an improved sense of privacy and independence for the patient. Write-N-Speak provided adults with aphasia with an extra level of stimulation through multimodal representations of language. The system also allowed some socially isolated individuals to participate in group activities, such as calling out Bingo numbers during game time.

In summary, this thesis investigated critical user groups and contexts for interaction, introduced novel systems to augment collocated communication, and looked closely at how these systems impact human-human interaction. While each project focused on a specific user group, the technical innovations and understanding of interaction has the potential to impact the broader population. Cooperative systems that support health care and therapy scenarios promise to play an important role in how people of all ages and abilities communicate about, manage, and share information.

Bibliography

Alač, M. and E. Hutchins (2004): I See What You are Saying: Action as Cognition in fMRI Brain Mapping Practice. *Cognition and Culture*, vol. 4.

Allen, M., J. McGrenere, and B. Purves (2008): The Field Evaluation of a Mobile Digital Image Communication Application Designed for People with Aphasia. *ACM Transactions on Accessible Computing*, vol. 1, no. 1, pp. 1–26.

Anoto (2011): Digital Pen and Paper Technology. http://www.anoto.com.

Argyle, M. and M. Cook (1976): *Gaze and mutual gaze*. Cambridge University Press.

Arias, E., H. Eden, and G. Fisher (1997): Enhancing communication, facilitating shared understanding, and creating better artifacts by integrating physical and computational media for design. In *DIS '97: Proceedings of the 2nd conference on Designing interactive systems*. New York, NY, USA, pp. 1–12, ACM.

Avery (2011): Office Products and Labels – Avery. http://www.avery.com.

Bachour, K., F. Kaplan, and P. Dillenbourg (2008): Reflect: An Interactive Table for Regulating Face-to-Face Collaborative Learning. In *EC-TEL '08: Proceedings of the 3rd European conference on Technology Enhanced Learning*. Berlin, Heidelberg, pp. 39–48, Springer-Verlag.

Baker, C. (1977): Regulators and turn-taking in American Sign Language discourse. In *On the other hand: New perspectives on American Sign Language*. New York, pp. 215–236, Academic Press.

Baker, C. and C. Padden (1978): Focusing on the nonmanual components of American Sign Language. In *Understanding Language through Sign Language Research*. New York, pp. 27–57, Academic Press.

Baron-Cohen, S. (2004): Mind Reading: The Interactive Guide to Emotions (CD-ROM).

Bauer, S. (1996): Asperger Syndrome. http://www.udel.edu/bkirby/asperger.

Beardon, L., S. Parsons, and H. Neale (2001): An interdisciplinary approach to investigating the use of virtual reality environments for people with Asperger Syndrome. *Educational and Child Psychology*, vol. 18.

Beattie, G. and H. Shovelton (1999): Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, vol. 123, no. 1-2, pp. 1.

Becvar, A., J. Hollan, and E. Hutchins (2005): Hands as Molecules: Representational Gestures for Developing Theory in Scientific Activity. *Semiotica*, vol. 156.

Bekker, M. M., J. S. Olson, and G. M. Olson (1995): Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *DIS '95: Proceedings of the 1st conference on Designing interactive systems*. New York, NY, USA, pp. 157–166, ACM.

Bernstein, M., A. Robinson-Mosher, R. B. Yeh, and S. R. Klemmer (2006): Diamond's Edge: From Notebook to Table and Back Again. In *UbiComp '06, Extended Abstracts*. Orange County, USA.

Bolt, R. A. (1980): "Put-that-there": Voice and gesture at the graphics interface. In *SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques*. New York, NY, USA, pp. 262–270, ACM.

Bungalow Software (2011): Therapy and Rehabilitation after stroke, aphasia and head injuries. http://www.bungalowsoftware.com.

Calder, D. (2008): SpeechKit: a multimedia speech tool. In *Proceedings of iiWAS '08*. Linz, Austria.

Cassell, J., D. McNeill, and K.-E. McCullough (1999): Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics cognition*, vol. 7, pp. 1.

Cavender, A., R. E. Ladner, and E. A. Riskin (2006): MobileASL: intelligibility of sign language video as constrained by mobile phone technology. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*. New York, NY, USA, pp. 71–78, ACM.

Chandler, S., J. Harris, A. Moncrief, and C. Lewis (2009): Naming practice for people with aphasia as a mobile web application. In *Proceedings of Assets '09*. Pittsburgh, USA.

Clark, A. (1997): *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.

Clark, H. (2003): Pointing and Placing. In S. Kita (ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Mihwah, NJ, pp. 243–268, Lawrence Erlbaum Associates.

Clark, H. and S. Brennan (1991): Grounding in Communication. In L. Resnick, J. Levine, and S. Teasley (eds.), *Perspectives on Socially Shared Cognition*. Washington, APA Books.

CloudGarden API (2008): CloudGarden Speech API. http://www.cloudgarden.com.

Cohen, J. (1960): A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*.

Cohen, P., C. Swindells, S. Oviatt, and A. Arthur (2008): A high-performance dual-wizard infrastructure for designing speech, pen, and multimodal interfaces. In *IMCI '08: Proceedings of the 10th international conference on Multimodal interfaces*. New York, NY, USA, pp. 137–140, ACM.

Cohen, P. R., M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, and J. Clow (1997): QuickSet: multimodal interaction for distributed applications. In *MULTIMEDIA '97: Proceedings of the fifth ACM international conference on Multimedia*. New York, NY, USA, pp. 31–40, ACM.

Cole, M. (1985): The Zone of Proximal Development: Where Culture and Cognition Create Each Other. In J. Wertsch (ed.), *Culture, Communication, and Cognition: Vygotskian Perspectives*. Cambridge, Cambridge University Press.

Corradini, A., R. Wesson, and P. Cohen (2002): A Map-Based System Using Speech and 3D Gestures for Pervasive Computing. In *ICMI '02: Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*. Washington, DC, USA, p. 191, IEEE Computer Society.

Devlin, S. and G. Unthank (2006): Helping aphasic people process online information. In *Proceedings of Assets '06*. Portland, USA.

Dietz, P. and D. Leigh (2001): DiamondTouch: a multi-user touch technology. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 219–226, ACM.

Duncan, S. (1972): Some signals and rules for taking turns in conversations. *Journal of personality and social psychology*, vol. 23, no. 2, pp. 283.

Duncan, S. (1974): On the Structure of Speaker-Auditor Interaction During Speaking Turns. *Language in society*, vol. 3, no. 2, pp. 161.

Duncan, S. and D. W. Fiske (1977): *Face-to-face interaction: Research, methods and theory*. Hilldale, NJ: Lawrence Erlbaum Associates.

Dynavox (2011): Speech Devices – Augmentative Communication. http://www.dynavoxtech.com.

Ehlers, S. and C. Gillberg (1993): The Epidemiology of Asperger's Syndrome: A Total Population Study. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 34.

Elwart-Keys, M., D. Halonen, M. Horton, R. Kass, and P. Scott (1990): User interface requirements for face to face groupware. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 295–301, ACM.

Emmorey, K., R. Thompson, and R. Colvin (2008): Eye Gaze During Comprehension of American Sign Language by Native and Beginning Signers. *Journal of Deaf Studies and Deaf Education*.

Engberg-Pedersen, E. (2003): From Pointing to Reference and Predication: Pointing Signs, Eyegaze, and Head and Body Orientation in Danish Sign Language. In S. Kita (ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum Associates.

Engeström, Y. (1990): Activity Theory and Individual and Social Transformation. In *International Congress for Research on Activity Theory, Lahti, Finland, May 21-25*.

Fairbanks, G. (1960): *Voice and Articulation Drillbook*. New York: Addison Wesley.

Fischer, G. and J. F. Sullivan (2002): Human-Centered Public Transportation Systems for Persons with Cognitive Disabilities - Challenges and Insights for Participatory Design. In *in 7th Participatory Design Conference*. pp. 194–198.

Garfinkel, H. (1967): *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Goffman, I. (1981): *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Goldin-Meadow, S. (2003): *Hearing Gesture: How our Hands Help Us Think*. Harvard University Press.

Goodwin, C. (1981): *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic Press.

Goodwin, C. (1986): Gestures as a resource for the organization of mutual orientation. *Semiotica*, vol. 62, no. 1/2, pp. 29.

Goodwin, C. (1995): Co-Constructing Meaning in Conversations with an Aphasic Man. *Research on Language in Social Interaction*, vol. 28.

Goodwin, C. (2000): Practices of Seeing, Visual Analysis: An Ethnomethodological Approach. In *Handbook of Visual Analysis*. London, pp. 157–182, Sage.

Goodwin, C. (2003a): Conversational Frameworks for the Accomplishment of Meaning in Aphasia. In C. Goodwin (ed.), *Conversation and Brain Damage*. Oxford: Oxford University Press.

Goodwin, C. (2003b): Pointing as Situated Practice. In S. Kita (ed.), *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ, Lawrence Erlbaum.

Goodwin, C. (2004): A Competent Speaker Who Can't Speak: The Social Life of Aphasia. *Journal of linguistic anthropology*, vol. 14, no. 2, pp. 151.

Goodwin, C. (2006): Human Sociality as Mutual Orientation in a Rich Interactive Environment: Multimodal Utterances and Pointing in Aphasia. In N. Enfield and S. C. Levinson (eds.), *Roots of Human Sociality*. London: Berg Press, pp. 96–125.

Goodwin, C. and M. Goodwin (1996): Formulating Planes: Seeing as Situated Activity. In *Cognition and Communication at Work*. pp. 61–95, Cambridge University Press.

Griffiths, M. (1997): Video games and clinical practice: issues, uses and treatments. *British Journal of Clinical Psychology*.

Guimbretière, F. (2003): Paper Augmented Digital Documents. In *Proceedings of UIST 2003, 16th Annual ACM Symposium on User Interface Software and Technology*. Vancouver, Canada, pp. 51–60.

Gullberg, M. and K. Holmqvist (2006): What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics cognition*, vol. 14, no. 1, pp. 53.

Hall, E. T. (1966): *The Hidden Dimension*. New York, NY, USA: New York: Anchor Books.

Haller, M., M. Billinghurst, J. Leithinger, D. Leitner, and T. Seifried (2005): Coeno: enhancing face-to-face collaboration. In *ICAT '05: Proceedings of the 2005 international conference on Augmented tele-existence*. New York, NY, USA, pp. 40–47, ACM.

Halloran, J., Y. Rogers, T. Rodden, and I. Taylor (2003): Creating new user experiences to enhance collaboration. In *INTERACT 2003, Zurich*.

Halverson, C. A. (2002): Activity Theory and Distributed Cognition: Or What Does CSCW Need to Do with Theories? *Computer Supported Cooperative Work*, vol. 11.

Han, J. Y. (2005): Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 115–118, ACM.

Heath, C. (1986): *Body movement and speech in medical interaction*. Cambridge University Press.

Heath, C. and P. Luff (1991): Collaborative activity and technological design: task coordination in London underground control rooms. In *Proceedings of the second conference on European Conference on Computer-Supported Cooperative Work*. Norwell, MA, USA, pp. 65–80, Kluwer Academic Publishers.

Heath, C. and P. Luff (1992): Collaboration and control: Crisis management and multimedia technology in London Underground line control rooms. *Computer Supported Cooperative Work*.

Hinckley, K., K. Yatani, M. Pahud, N. Coddington, J. Rodenhouse, A. Wilson, H. Benko, and B. Buxton (2010): Pen + touch = new tools. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 27–36, ACM.

Hindmarsh, J. and C. Heath (2000): Embodied reference: A study of deixis in workplace interaction. *Journal of Pragmatics*, vol. 32, no. 12, pp. 1855.

Hollan, J., E. Hutchins, and D. Kirsh (2000): Distributed Cognition: Toward a New Theoretical Foundation for Human-Computer Interaction Research. *ACM Transactions on Human-Computer Interaction*, pp. 174–196.

Hollan, J. D. and E. L. Hutchins (2009): Opportunities and Challenges for Augmented Environments: A Distributed Cognition Perspective. In S. Lahlou (ed.), *Designing User Friendly Augmented Work Environments: From Meeting Rooms to Digital Collaborative Spaces*. Springer.

Hutchins, E. (1995): *Cognition in the Wild*. Cambridge, MA: MIT Press.

Hutchins, E. and L. Palen (1997): Constructing Meaning from Space, Gesture, and Speech. In L. B. Resnick, R. Saljo, C. Pontecorvo, and B. Burge (eds.), *Discourse, tools, and reasoning: Essays on situated cognition*. Heidelberg, Germany, Springer-Verlag.

iCommunicator (2008): www.myicommunicator.com.

Inkpen, K., K. S. Booth, S. D. Gribble, and M. Klawe (1995): Give and take: children collaborating on one computer. In *Conference companion on Human factors in computing systems*. New York, NY, USA, pp. 258–259, ACM.

Inkpen, K., J. McGrenere, K. S. Booth, and M. Klawe (1997): The effect of turn-taking protocols on children's learning in mouse-driven collaborative environments. In *Proceedings of the conference on Graphics interface '97*. Toronto, Ont., Canada, Canada, pp. 138–145, Canadian Information Processing Society.

Java Speech API (2008): java.sun.com/products/java-media/speech.

Jefferson, G. (2004): Glossary of Transcript Symbols with an Introduction. In *Conversation Analysis: Studies from the First Generation*. Jesse Benjamins Publishing Company.

Jiang, H., R. B. Yeh, T. Winograd, and Y. Shi (2007): DigiPost: Writing on Post-its with Digital Pen to Support Collaborative Editing Tasks on Tabletop Displays. In *UIST '07 Posters*.

Johnson, M. (1987): *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University of Chicago Press.

Kendon, A. (1967): Some functions of gaze-direction in social interaction. *Acta Psychologica*, vol. 26, no. 1, pp. 22–63.

Kendon, A. (1980): Gesticulation and Speech: Two Aspects of the Process of Utterance. In *The Relationship of Verbal and Nonverbal Communication*. p. 388, Walter de Gruyter.

Kendon, A. and C. Muller (2001): Introducing: GESTURE. *Gesture*, vol. 1, no. 1, pp. 1.

Kerr, S. J., H. R. Neale, and S. V. G. Cobb (2002): Virtual environments for social skills training: the importance of scaffolding in practice. In *Proceedings of the fifth international ACM conference on Assistive technologies*. New York, NY, USA, pp. 104–110, ACM.

Kientz, J., S. Boring, G. Abowd, and G. Hayes (2005): Abaris: Evaluating Automated Capture Applied to Structured Autism Interventions. In *Proceedings of Ubicomp '05*. Tokyo, Japan.

Kirk, D., A. Crabtree, and T. Rodden (2005): Ways of the hands. In *ECSCW'05: Proceedings of the ninth conference on European Conference on Computer Supported Cooperative Work*. New York, NY, USA, pp. 1–21, Springer-Verlag New York, Inc.

Kita, S. (2003): *Pointing: where language, culture, and cognition meet*. Lawrence Erlbaum Associates.

Kleinke, C. (1986): Gaze and eye contact: A research review. *Psychological Bulletin*, vol. 100, no. 1, pp. 78–100.

Konishi, K. and N. F. Ikeda (2007): Data Model and Architecture of a Paper-Digital Document Management System. In *Proceedings of DocEng '07, ACM symposium on Document Engineering*. Winnipeg, Canada, pp. 29–31.

Kraut, R. E., S. R. Fussell, and J. Siegel (2003): Visual information as a conversational resource in collaborative physical tasks. *Hum.-Comput. Interact.*, vol. 18, no. 1, pp. 13–49.

Krueger, M. W. (1990): VIDEOPLACE and the Interface of the Future. In B. Laurel (ed.), *The Art of Human Interface Design*. Addison-Wesley.

Krueger, M. W. (1991): *Artificial Reality II*. Reading, MA: Addison-Wesley.

Krueger, M. W. (1993): Environmental technology: making the real world virtual. *Commun. ACM*, vol. 36, pp. 36–37.

Krueger, M. W., T. Gionfriddo, and K. Hinrichsen (1985): VIDEOPLACE–an artificial reality. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 35–40, ACM.

Kruger, R., S. Carpendale, S. D. Scott, and S. Greenberg (2003): How people use orientation on tables: comprehension, coordination and communication. In *GROUP '03: Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*. New York, NY, USA, pp. 369–378, ACM.

Leont'ev, A. (1978): *Activity, Consciousness, and Personality*. Englewood Cliffs, NJ: Prentice-Hall.

Liao, C., F. Guimbretière, K. Hinckley, and J. Hollan (2008): PapierCraft: A Gesture-Based Command System for Interactive Paper. *ACM Transactions on Computer Human Interaction*, vol. 14, no. 4, pp. 1–27.

Liddell, S. K. and M. Metzger (1998): Gesture in sign language discourse. *Journal of Pragmatics*, vol. 30, no. 6, pp. 657 – 697.

Lingraphica (2011): Communication Device for Stroke Victims – Aphasia – Apraxia. http://www.aphasia.com.

Livescribe (2011): Livescribe smart pen technology. http://www.livescribe.com.

Luff, P. and C. Heath (2000): The collaborative production of computer commands in command and control. *Int. J. Hum.-Comput. Stud.*, vol. 52, pp. 669–699.

Luff, P., C. Heath, and D. Greatbatch (1992): Tasks-in-interaction: paper and screen based documentation in collaborative activity. In *Proc. CSCW '92*. pp. 163–170.

Luff, P., C. Heath, H. Kuzuoka, K. Yamazaki, and J. Yamashita (2006): Handling documents and discriminating objects in hybrid spaces. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York, NY, USA, pp. 561–570, ACM.

MacKay, W. (1999): Is paper safer? The role of paper flight strips in air traffic control. *ACM TOCHI*, vol. 6, no. 4, pp. 340.

Mackay, W. E., G. Pothier, C. Letondal, K. Bøegh, and H. E. Sørensen (2002): The Missing Link: Augmenting Biology Laboratory Notebooks. In *Proc. UIST '02*. Paris, France, pp. 41–50.

Magerkurth, C., M. Memisoglu, T. Engelke, and N. Streitz (2004): Towards the next generation of tabletop gaming experiences. In *GI '04: Proceedings of Graphics Interface 2004*. School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, pp. 73–80, Canadian Human-Computer Communications Society.

Mandryk, R. L. and D. S. Maranan (2002): False prophets: exploring hybrid board/video games. In *CHI '02 extended abstracts on Human factors in computing systems*. New York, NY, USA, pp. 640–641, ACM.

Mandviwalla, M. and L. Olfman (1994): What do groups need? A proposed set of generic groupware requirements. *ACM Trans. Comput.-Hum. Interact.*, vol. 1, no. 3, pp. 245–268.

Matsushita, M., M. Iida, T. Ohguro, Y. Shirai, Y. Kakehi, and T. Naemura (2004): Lumisight table: a face-to-face collaboration support system that optimizes direction of projected information to each stakeholder. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*. New York, NY, USA, pp. 274–283, ACM.

Matthews, T., S. Carter, C. Pai, J. Fong, and J. Mankoff (2006): Scribe4me: Evaluating a mobile sound translation tool for the deaf. In *Proceedings of Ubicomp 2006 International Conference on Ubiquitous Computing*. pp. 159–176, Springer.

Matthews, T., J. Fong, and J. Mankoff (2005): Visualizing non-speech sounds for the deaf. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*. New York, NY, USA, pp. 52–59, ACM.

McNeill, D. (1985): So You Think Gestures Are Nonverbal? *Psychological review*, vol. 92, no. 3, pp. 350.

McNeill, D. (1992): *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

Miller, D., K. Gyllstrom, D. Stotts, and J. Culp (2007): Semi-transparent video interfaces to assist deaf persons in meetings. In *ACM-SE 45: Proceedings of the 45th annual southeast regional conference*. New York, NY, USA, pp. 501–506, ACM.

Moffatt, K., J. McGrenere, B. Purves, and M. Klawe (2004): The participatory design of a sound and image enhanced daily planner for people with aphasia. In *Proceedings of CHI '04*. Vienna, Austria.

Morris, M. R. (2006): Supporting Effective Interaction with Tabletop Groupware. Ph.D. thesis, Stanford University, Stanford, CA.

Morris, M. R., A. Huang, A. Paepcke, and T. Winograd (2006): Cooperative gestures: multi-user gestural interactions for co-located groupware. In *Proc. of CHI '06*. New York: ACM Press, 1201-1210.

Morris, M. R., D. Morris, and T. Winograd (2004a): Individual audio channels with single display groupware: effects on communication and task strategy. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. New York, NY, USA, pp. 242–251, ACM.

Morris, M. R., K. Ryall, C. Shen, C. Forlines, and F. Vernier (2004b): Beyond "social protocols": multi-user coordination policies for co-located groupware. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. New York, NY, USA, pp. 262–265, ACM.

Nardi, B. A. (1996): Studying context: A comparison of activity theory, situated action models, and distributed cognition. In *Context and Consciousness: Activity Theory and Human-Computer Interaction*.

National Stroke Association (2010): Stroke 101 Fact Sheet. http://www.stroke.org.

NIDCD (2008): National Institute on Deafness and Communication Disorders.

Nomura, S., E. Hutchins, and B. Holder (2006): The uses of paper in commercial airline flight operations. In *Proc. CSCW '06*. pp. 249–258.

Norman, D. (1993): *Things That Make Us Smart*. Reading, MA: Addison-Wesley Publishing Company.

Norrie, M. C., B. Signer, and N. Weibel (2006): General Framework for the Rapid Development of Interactive Paper Applications. In *Proceedings of CoPADD 2006, 1st International Workshop on Collaborating over Paper and Digital Documents*. Banff, Canada, pp. 9–12.

Nunez, R. (1999): Could the Future Taste Purple? Reclaiming Mind, Body, and Cognition. In R. Núñez and W. Freeman (eds.), *Reclaiming Cognition: The Primacy of Action, Intention and Emotion*. Thorverton, UK, Imprint Academic.

Oviatt, S. and R. VanGent (1996): Error resolution during multimodal human-computer interaction. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, Vol. 1. pp. 204 –207 vol.1.

Oviatt, S. L. (1999): Ten Myths of Multimodal Interaction. *Communications of the ACM*, vol. 42.

Oviatt, S. L. (2002): Breaking the Robustness Barrier: Recent Progress on the Design of Robust Multimodal Systems. *Advances in Computers*, vol. 56.

Pangaro, G., D. Maynes-Aminzade, and H. Ishii (2002): The actuated workbench: computer-controlled actuation in tabletop tangible interfaces. In *UIST '02: Proceedings of the 15th annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 181–190, ACM.

Parker, J. K., R. L. Mandryk, and K. M. Inkpen (2005): TractorBeam: seamless integration of local and remote pointing for tabletop displays. In *GI '05: Proceedings of Graphics Interface 2005*. School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, pp. 33–40, Canadian Human-Computer Communications Society.

Patten, J., H. Ishii, J. Hines, and G. Pangaro (2001): Sensetable: a wireless object tracking platform for tangible user interfaces. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 253–260, ACM.

Rekimoto, J. (2002): SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 113–120, ACM.

Rekimoto, J. and M. Saitoh (1999): Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*. New York, NY, USA, pp. 378–385, ACM.

Ringel, M., K. Ryall, C. Shen, C. Forlines, and F. Vernier (2004a): Release, relocate, reorient, resize: fluid techniques for document sharing on multi-user interactive tables. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*. New York, NY, USA, pp. 1441–1444, ACM.

Ringel, M., K. Ryall, C. Shen, C. Forlines, and F. Vernier (2004b): Release, relocate, reorient, resize: fluid techniques for document sharing on multi-user interactive tables. In *CHI '04 extended abstracts on Human factors in computing systems*. New York, NY, USA, pp. 1441–1444, ACM.

Rodden, T., Y. Rogers, J. Halloran, and I. Taylor (2003): Designing novel interactional workspaces to support face to face consultations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 57–64, ACM.

Rogers, Y. and J. Ellis (1994): Distributed Cognition: An alternative framework for analysing and explaining collaborative working. *Journal of Information Technology*.

Rogers, Y., W. Hazlewood, E. Blevis, and Y.-K. Lim (2004): Finger talk: collaborative decision-making using talk and fingertip interaction around a tabletop display. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*. New York, NY, USA, pp. 1271–1274, ACM.

Rogers, Y. and S. Lindley (2004): Collaborating around vertical and horizontal large interactive displays: which way is best? *Interacting with Computers*, vol. 16, no. 6, pp. 1133 – 1152.

Rogers, Y. and T. Rodden (2003): Configuring spaces and surfaces to support collaborative interactions. In K. O'Hara, M. Perry, E. Churchill, and D. Russell (eds.), *Public and Situated Displays*. Kluwer Publishers.

Roth, W.-M. (2001): Gestures: Their Role in Teaching and Learning. *Review of Educational Research*, vol. 71, no. 3, pp. 365–392.

Roth, W.-M. and D. V. Lawless (2002): When up is down and down is up: Body orientation, proximity, and gestures as resources. *Language in Society*, vol. 31, no. 01, pp. 1–28.

Salomon, G. (1993): *Distributed Cognitions*. Cambridge University Press.

Schegloff, E. A. (1984): On Some Gestures' Relation to Talk. In J. M. Atkinson and J. Heritage (eds.), *Structures of Social Action*. Cambridge: Cambridge University Press.

Schegloff, E. A. and H. Sacks (1973): Opening Up Closings. *Semiotica*, vol. 8.

Schuler, D. and A. Namioka (eds.) (1993): *Participatory Design: Principles and Practices*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc.

Schull, J. (2006): An extensible, scalable browser-based architecture for synchronous and asynchronous communication and collaboration systems for deaf and hearing individuals. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*. New York, NY, USA, pp. 285–286, ACM.

Scott, S. D., M. Sheelagh, T. Carpendale, and K. M. Inkpen (2004): Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. New York, NY, USA, pp. 294–303, ACM.

Sellen, A. J. and R. Harper (2001): *The Myth of the Paperless Office*. MIT Press.

Shen, C., K. Everitt, and K. Ryall (2003): UbiTable: Impromptu Face-to-Face Collaboration on Horizontal Interactive Surfaces. In *In Proc. UbiComp 2003*. pp. 281–288.

Shen, C., F. D. Vernier, C. Forlines, and M. Ringel (2004): DiamondSpin: an extensible toolkit for around-the-table interaction. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 167–174, ACM.

Signer, B. (2008): *Fundamental Concepts for Interactive Paper and Cross-Media Information Spaces*. Books on Demand Gmbh.

Signer, B., U. Kurmann, and M. C. Norrie (2007): iGesture: A General Gesture Recognition Framework. In *Proceedings of ICDAR 2007, 9th International Conference on Document Analysis and Recognition*. Curitiba, Brazil.

Signer, B. and M. C. Norrie (2007): PaperPoint: a paper-based presentation and interactive paper prototyping tool. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. New York, NY, USA, pp. 57–64, ACM.

Siple, P. (1978): Visual Constraints for Sign Language Communication. *Sign Language Studies*.

Starner, T., J. Weaver, and A. Pentland (1997): A Wearable Computer Based American Sign Language Recognizer. In *ISWC'97*. pp. 130–137.

Steimle, J., O. Brdiczka, and M. Mühlhäuser (2009): CoScribe: Integrating Paper and Digital Documents for Collaborative Kknowledge Work. *IEEE Transactions on Learning Technologies*, vol. 2, no. 3, pp. 174–188.

Stewart, J., B. B. Bederson, and A. Druin (1999): Single display groupware: a model for co-present collaboration. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 286–293, ACM.

Stifelman, L., B. Arons, and C. Schmandt (2001): The Audio Notebook: Paper and Pen Interaction with Structured Speech. In *Proc. CHI '01*. Seattle, WA, USA, pp. 182–189.

Suchman, L. (1987): *Plans and situated actions : The Problem of Human-Machine Communication*. New York: Cambridge University Press.

Sugimoto, M., K. Hosoi, and H. Hashizume (2004): Caretta: a system for supporting face-to-face collaboration by integrating personal and shared spaces. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 41–48, ACM.

Tandler, P., T. Prante, C. Müller-Tomfelde, N. Streitz, and R. Steinmetz (2001): Connectables: dynamic coupling of displays for the flexible creation of shared workspaces. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 11–20, ACM.

Tang, J. C. (1991): Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, vol. 34, no. 2, pp. 143 – 160. Special Issue: Computer-supported Cooperative Work and Groupware. Part 1.

Tang, J. C. and S. L. Minneman (1990): VideoDraw: a video interface for collaborative drawing. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 313–320, ACM.

Team Asperger (2004): Gaining Face: Facial recognition software. http://ccoder.com/GainingFace.

Tee, K., K. Moffatt, L. Findlater, E. MacGregor, J. McGrenere, B. Purves, and S. S. Fels (2005): A visual recipe book for persons with language impairments. In *Proceedings of CHI '05*. Portland, USA.

Terrenghi, L., D. Kirk, A. Sellen, and S. Izadi (2007): Affordances for manipulation of physical versus digital media on interactive surfaces. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 1157–1166, ACM.

The National Aphasia Association (2011): http://www.aphasia.org.

Touchlib (2011): A Multitouch Development Kit. http://nuigroup.com/touchlib.

Tse, E. (2007): Multimodal Co-Located Interaction. Ph.D. thesis, The University of Calgary, Calgary, Alberta, Canada.

Tse, E., S. Greenberg, C. Shen, and C. Forlines (2007a): Multimodal multiplayer tabletop gaming. *Comput. Entertain.*, vol. 5, no. 2, pp. 12.

Tse, E., S. Greenberg, C. Shen, C. Forlines, and R. Kodama (2008): Exploring true multi-user multimodal interaction over a digital table. In *DIS '08: Proceedings of the 7th ACM conference on Designing interactive systems*. New York, NY, USA, pp. 109–118, ACM.

Tse, E., C. Shen, S. Greenberg, and C. Forlines (2006): Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. In *Proceedings of the working conference on Advanced visual interfaces*. New York, NY, USA, pp. 336–343, ACM.

Tse, E., C. Shen, S. Greenberg, and C. Forlines (2007b): How pairs interact over a multimodal digital table. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 215–218, ACM.

Turoff, M. (1975): Computerized conferencing for the deaf and handicapped. *SIGCAPH Comput. Phys. Handicap.*, pp. 4–11.

Ullmer, B. and H. Ishii (1997): The metaDESK: models and prototypes for tangible user interfaces. In *UIST '97: Proceedings of the 10th annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 223–232, ACM.

Varela, F., E. Thompson, and E. Rosch (1991): *The Embodied Mind*. Cambridge: MIT Press.

Vygotsky, L. (1978): *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weibel, N. (2009): A Publishing Infrastructure for Interactive Paper Documents: Supporting Interactions across the Paper-Digital Divide. Ph.D. thesis, ETH Zurich.

Weibel, N., A. Ispas, B. Signer, and M. C. Norrie (2008): Paperproof: a paper-digital proof-editing system. In *CHI '08 extended abstracts on Human factors in computing systems*. New York, NY, USA, pp. 2349–2354, ACM.

Weiss, M., J. Wagner, Y. Jansen, R. Jennings, R. Khoshabeh, J. D. Hollan, and J. Borchers (2009): SLAP widgets: bridging the gap between virtual and physical controls on tabletops. In *Proceedings of CHI '09*. New York: ACM Press, 481-490.

Wellner, P. (1991): The DigitalDesk calculator: tangible manipulation on a desk top display. In *UIST '91: Proceedings of the 4th annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 27–33, ACM.

Wellner, P. (1993): Interacting with paper on the DigitalDesk. *Communications of the ACM*, vol. 36, no. 7, pp. 87 – 96.

Wertsch, J. (1981): *The Concept of Activity in Soviet Psychology*. Armonk, NY: M. E. Sharpe.

Whalen, T. (2003): Playing Well with Others: Applying Board Game Design to Tabletop Display Interfaces. In *UIST '03: Extended proceedings of the 2003 international conference on User Interface Software and Technology*. New York, NY, USA, ACM.

Wilson, A. D. (2004): TouchLight: an imaging touch screen and display for gesture-based interaction. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*. New York, NY, USA, pp. 69–76, ACM.

World Health Organization (2007): World Health Organization. *World Health Report*. http://www.who.int/whr/en.

World Health Organization (2011a): Ageing and Life Course. http://www.who.int/ageing/en.

World Health Organization (2011b): Disabilities. http://www.who.int/topics/disabilities/en/.

Wu, M., R. Baecker, and B. Richards (2005): Participatory design of an orientation aid for amnesics. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA, pp. 511–520, ACM.

Wu, M. and R. Balakrishnan (2003): Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*. New York, NY, USA, pp. 193–202, ACM.

Yeh, R., C. Liao, S. Klemmer, F. Guimbretière, B. Lee, B. Kakaradov, J. Stamberger, and A. Paepcke (2006): ButterflyNet: a mobile capture and access system for field biology research. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York, NY, USA, pp. 571–580, ACM.

Yeh, R. B., S. R. Klemmer, A. Paepcke, M. Bastéa-Forte, J. Brandt, and J. Boli (2007): Iterative Design of a Paper + Digital Toolkit: Supporting Designing, Developing, and Debugging. Technical Report CSTR 2007-10, Stanford University, Computer Science Department.

Zemel, A., T. Koschmann, C. Lebaron, and P. Feltovich (2008): "What are We Missing?" Usability's Indexical Ground. *Computer Supported Cooperative Work*, vol. 17, no. 1, pp. 63–85.