Display Design Trade-offs for a Wireless Mobile Captioning System

Leanne L. West[1], Ethan W. Adler[1],
Jeff Jo[1], John. M. Stewart[1], and Jack W. Wood[1]

[1] Georgia Tech Research Institute, Electro-Optical Systems Laboratory, 925 Dalney Street,
Atlanta, GA 30332, USA
{leanne.west, ethan.adler, jeff.jo, john.stewart, jack.wood}@gtri.gatech.edu

Abstract. There are over 31 million people in the United States with hearing loss, and this number is expected to grow by to 41 million people by 2025. To be completely socially responsible, mobile computing needs to be accessible by all persons, including those with disabilities. Additionally, these platforms can help address the unmet needs of people with disabilities to make their everyday lives easier, more complete, and more rewarding. This paper will follow the development of a mobile wireless captioning system, including human subjects testing and design tradeoffs that focus on a microdisplay for the system.

Keywords: wireless, mobile, captioning, deaf, hard of hearing, microdisplay, smartphone, PDA.

1 Introduction

According to Kochkin (2004), there are over 31 million people in the United States with hearing loss, with baby boomers and the 75 and older age brackets seeing the largest increase between 1989 and 2004. This number is expected to grow by another third to 41 million people by 2025.[1,2] Of the 31 million who have hearing loss currently, approximately 1 million are functionally deaf [3], while the remainder are hard of hearing.

Researchers that the Georgia Tech Research Institute (GTRI) developed a mobile wireless personal captioning system to bring captions to the millions of people who patronize venues such as movie theaters, sports arenas, places of worship, and schools, who are unable to fully experience these events because they cannot hear the information that is being presented audibly at the venue. The lack of captioning in community venues prevents equal access to information for individuals with hearing impairments and prohibits their full involvement in the community. Current forms of text presentation include large communal screens or systems that are specific to only one venue, such as movie theaters. Additional uses for the system include the transmission of multiple languages and the transmission of other information that might be desired, like statistics at a sporting event.

Because the captioning system, or Communication Assistant, is intended for real-world use, the system has been tested with end users to obtain feedback on usability and user interface issues. Usability testing has focused on the microdisplay needed for the system, which led to some obvious and some not-so-obvious results. Following an overview of the captioning system as a whole, this paper will focus on user testing and design trade-offs associated with the microdisplay portion of the system.

2 The Communication Assistant System Overview

The Communication Assistant uses the 802.11b wireless protocol and consists of three components: a transmitter, a receiver, and a display. Display technology includes wirelessly-enabled smartphones, PDAs (Personal Digital Assistants), and laptop computers. For some applications, though, users may require a display that is smaller, more private, and hands-free. For those instances, a miniature display is worn in front of one or both eyes. This display is plugged into a PDA or smartphone. The smartphone functions as the wireless receiver that controls the microdisplay and provides power.

Although positioned close to the eye, the microdisplay uses optics that make its screen appear to float several feet away, giving users relaxed viewing of both the captions and the world around them. The display, shown in Figure 1(a), creates the illusion of captioned text overlaid on the user's visual field; the wearer’s real-world view is augmented by the captions, as demonstrated in Figure 1(b).

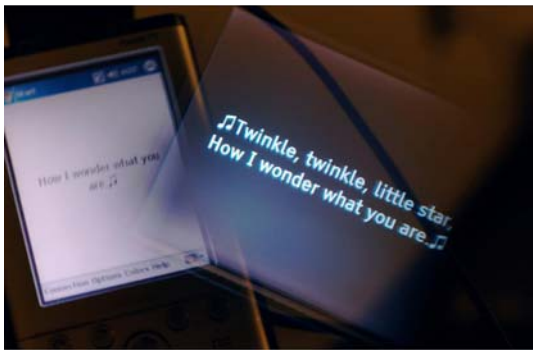

Figure 1(a) Captioning system (b) Microdisplay text "floating" in front of PDA screen.

Input for the transmitter can be pre-recorded, typed in real-time, or generated by speech recognition. In cases where captions can be created ahead of time, they will be stored digitally then transferred to the transmitter. There are two methods of input allowed for captions that are generated in real time. One method is CART (Communication Access Real-time Translation or Computer Assisted Real-time Translation), and the other is automatic speech recognition (ASR) technology. A CART provider transcribes a speaker's comments verbatim using a stenograph keyboard and specialized software. The output from the stenograph station is delivered directly into the Communication Assistant transmitter. CART is the current preferred method to generate captions and transcripts. Speech recognition technology in its current state has several problems. A recent analysis of performance trends in computer speech recognition systems estimated that human-like speech recognition performance would not be achieved until somewhere around 2040 or 2050.[4] Even before ASR reaches human-like performance though, it will prove to be a beneficial tool for providing captions, especially when punctuation issues are resolved.

In some environments such as movie theaters where there is the need for multiple simultaneous data streams, the wireless system will utilize multiple transmission "channels". "Channel" options give the advantage of using the same hardware to access multiple streams of information in a single facility. The simultaneous availability of different text streams also allows the accommodation of people who do not speak English or who use English as a second language by providing the possibility to send multiple languages at the same time.

Channel options are just one form of system control given to users. The user interface also provides options for text size; text and background colors; pop-up or scrolling captions; monocular or binocular versions of the microdisplay; quick inversion of text/background color; focus adjustment, and center or left justification of the text. Additionally, the smartphone and microdisplay can be formatted independently from each other. The formatting options stemmed from the results of the user testing, to be discussed in the following section.

3 Design Considerations for the Display

There are many considerations that go into the design of any system. Sometimes it is difficult to balance the desired features of the end user with what is feasible with today's technology. Trade-offs always exist between size, weight, cost, functionality, ease of use, and practicality for the real-world application. In the end, it is important to consider the options and weigh the benefits and disadvantages that each component brings to the system, while keeping in mind that the most important person to consider is the end user. In addition, while one display may be best for a certain use or target population, it does not mean that it is correct for all uses and users.

There are several options for the display component of the Communication Assistant system. Currently, the end user is able to watch captions on the screen of the smartphone, on a laptop screen, or on a microdisplay screen. Different situations or venues warrant the use of the different displays, which is why flexibility was built into the system. In some venues, like a place of worship or a conference meeting, the smartphone screen might offer the best viewing option. In a classroom setting, where computers can be a common feature, the computer monitor may provide the best option for caption display. In this age of multi-media presentations, some classroom lectures could also benefit from a head-mounted display or microdisplay. Microdisplays are also necessary in venues such as movie theaters, live theater, and sports arenas where a large focus of the event is on visual action.

There are several different types and styles of microdisplays. Each different microdisplay has various features and characteristics that were considered when choosing a microdisplay for this system. Many head-mounted displays are not "micro" at all; they are very large with almost a "helmet-like" appearance. These systems are not discreet or

lightweight, and are therefore not the choice for a system that will be utilized for multiple hours in a row by a group of people who would rather not have it known that they need a special system for assistance.

Some head-mounted displays employ retinal writing, which creates the image with the use of a laser, usually red in color. The red color of the laser limits the customizability of the system, and studies have shown [5] that white text is preferred over colored text for captions. Another issue with this type of display has to do with marketability of the commercial product. During our testing (discussed in the next section), researchers learned that many people who have a hearing impairment are very protective of their eyesight. While these retinal writing displays are not at all dangerous, the concept of using a laser to create the image can be an unsettling thought for the consumer.

4 User Input and Testing

The project began by approaching a local advocacy group for people who are deaf and hard of hearing. The idea for the project was posed to the Georgia Council for the Hearing Impaired (GaCHI) to make sure that the envisioned system would be of interest to the population it was intended to serve. It is imperative to involve members of the population of interest from the very beginning when creating a technology that serves a specific population. They will know if they are likely to use what is being proposed for development, and they will have insight as to what features are important for them.

During the course of the Communication Assistant project, two formal rounds of laboratory testing occurred to elicit feedback and comments from the deaf and hard of hearing community. GaCHI played a vital role in recruiting volunteers for the initial round of testing. The device was further tested at the Self Help for the Hard of Hearing (now called the Hearing Loss Association of America) annual conference. The tests were conducted in a simulated movie-theater setting with one to four volunteers at a time. After a brief introduction to the device, volunteers were asked to watch anywhere from 15 minutes of a movie to a full-length feature (up to 1 1/2 hours), while using a monocular microdisplay system to receive the captioning of the movie. The following information and quotes come directly from the questionnaires that the volunteers filled out after using the device to watch the movie.

Including both rounds of testing, 80 participants were recruited to test the device, representing a variety of age groups ranging from 15 to over 75 years old. Volunteers included people who are hearing (usually an ASL interpreter or an interested family member), deaf from birth, late deafened, with the majority of testers being hard of hearing. Almost all of the adult volunteers had a college degree or higher.

As anticipated, most volunteers thought the monocular display "took some getting used to," but it normally took less than 10 minutes to get the display in comfortable reading position. Most volunteers also thought text was easy to read, but desired a choice of fonts. The most noted suggestions to improve the readability were to include a text-size option and a display focus adjustment, both of which were not part of the original prototypes but have been included in the final product. The focus adjustment allows the wearer to focus the text at a desired distance, for example in the plane of the movie screen. This adjustment relieves any strain caused by a difference in depth perception of the words and the object or scene being viewed. From the researchers' own experiences, wearing a microdisplay of any type is different and requires some adaptation. However, after several uses of the device, the wearer becomes accustomed to the microdisplay and adaptation is almost instantaneous. The researchers have found that adaptation is even easier when wearing the binocular system.

The desire for custom formatting of text was strongly indicated during testing, including text size, text and background color, justification of the text, etc. Many of the younger testers wanted to have multiple options for background and text color, while more experienced readers of captions preferred a black background with white text. The more experienced caption readers also preferred pop-up style captions in two lines of text because this is the format in which they are used to receiving captions. However, slower readers preferred to keep a full screen of text in case they fell behind or wanted to look back at what had been said. The ability to customize the captions to suit the individual is not met by current text presentation systems available in some movie theaters or where words are displayed on a communal screen available in some live theater, classrooms, and places of worship.

Almost none of the volunteers thought that wearing the microdisplay would make them feel self-conscious. The few that said it would gave comments such as: "wouldn't mind in theater, but in lighted situation would be hesitant unless with others wearing it too."; "I am only self conscious a little. I would use it in a movie theater no problem. I might be more self conscious when lights are on."; "Only if I'm the only one using it."; "If it helps enough for me to get what is going on in the movie then I don't care what others think." Since this testing occurred, the electronic eyewear has improved in fit, comfort, usability, and style and more and more mobile wireless technologies have

come on the market such as the Microsoft Zune and Apple Video iPOD, making head-mounted electronics even more common and ubiquitous.

The system has been demonstrated and tested in actual movie theaters as well. Volunteers were asked to watch a 20 or 40 minute segment of a movie with the monocular system and then a heavier and bulkier binocular system. Even though all previous testing had indicated that smaller, lighter, and more discreet were the desired specifications of the head-mounted display, actual comparison between the two systems yielded a surprising result. Movie-going volunteers liked the binocular version better than the monocular version because use of a binocular version was easier. The binocular version goes on like a pair of glasses, while the monocular version has one to five minutes of alignment the first time it is worn. Ease of use outweighed the parameters such as weight, size, and discreetness.

The system has also been demonstrated with a member of the deaf community at the Georgia Tech baseball stadium. Pictures from the event are shown in Figures 2. A CART reporter provided the transcription of the announcer and color commentary of the game. During the demonstration, the first question asked by the deaf volunteer was, "Where do I have to sit?" When he was told that he could sit anywhere (an unusual experience for someone who needs text captions or an American Sign Language interpreter), he was very excited and chose to sit at the far end of the stadium near the left fielder. Overall, he was happy with the system and felt that it provided him a more complete experience at the ball game.

Figure 2. Images from the baseball game demonstration, held at the Georgia Tech Baseball Stadium.

5 Conclusions

As stated previously, there are over 30 million people who are hard of hearing, and this number is growing every day as the population ages. There is a social responsibility to use technology to allow all people full involvement in their community. The mobile captioning system discussed here is an example of how mobile technologies can enrich the everyday lives of individuals with special needs. Proof-of-concept prototypes of the Communication Assistant have been built and tested within the deaf and hard of hearing communities. Real-world testing and demonstrations have been held in movie theaters and at a baseball stadium. Feedback on the system, while sometimes yielding unexpected results, has been extremely positive, and a commercial product is currently being finalized with an industry partner.

Even when you define requirements for a system that will be used "in the wild" with the individuals who will use the system, their ideas and wants on paper may not actually be what they want. This fact, coupled with all other design trade-offs necessary to build a system can make the job of an engineer very tricky.

Acknowledgements

Funding for the Communication Assistant was provided by the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education under grant number H133E010804 and the Georgia Tech Research Institute. The opinions contained in this publication are those of the grantee and do not necessarily reflect those of the U.S. Department of Education or the Georgia Tech Research Institute.

References

1. Kochkin S. MarkeTrakVI: The VA and direct mail sales spark growth in hearing aid market. The Hearing Review December 2001; 8(12): 16-24, 63-65.
2. Kochkin S. World Age Demography Data-base. Better Hearing Institute August 2004.
3. Mitchell R. How Many Deaf People Are There in the United States? Estimates From the Survey of Income and Program Participation. *The Journal of Deaf Studies and Deaf Education* 2006; 11(1):112-119.
4. Arrowood J. State of The Art in Speech Synthesis, Recognition, and Understanding. *Wireless RERC State of Technology Conference*, May 2004.
5. National Center for Accessible Media. Advanced Television Closed Captioning Research Report. Online posting date: 1998. Available at http://www.cfv.org/caai/nadh128.pdf. Accessed 2006 Feb 14**.**