THESIS

SOCIALEYES: DEVELOPING A USEFUL INTERFACE FOR THE VISUALLY DISABLED

Submitted by

Erin Nagoshi

Department of Computer Science

In partial fulfillment of the requirements

for the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2011

Master's Committee:

Advisor: Bruce Draper

Ross Beveridge
Lucy Troup

ABSTRACT

SOCIALEYES: DEVELOPING A USEFUL INTERFACE FOR THE VISUALLY DISABLED

While many tools exist to help the visually disabled navigate, there are very few designed for social situations. Recent advancements in the field of facial recognition offer the opportunity to change that. This thesis begins a study of the human computer interaction challenges of developing usable interfaces for visual social aides.

ACKNOWLEDGEMENTS

Thank you to volunteers Christopher Cambell, Christopher Millard, Benjamin Say, Brandon Schaffer, Craig Shaffer, Steven Wiersma, and Maggie Wigness for their support and willingness to participate in this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Being blind can be a lonely ordeal. Sighted people often feel uncomfortable around the visually disabled because they don't know how to act. In order to abstain from saying or doing anything offensive, they sometimes avoid interaction with the blind altogether. Many blind people are able to adapt to this kind of isolation by developing extrovert personalities and taking it upon themselves to teach sighted people how to feel comfortable around them. Those who are naturally shy, however, have a much more difficult time meeting people.

From the perspective of a blind person, social situations such as small parties or gatherings, can be difficult. Locating someone you know can be challenging when you can't see them. Sometimes the visually disabled can locate and identify people using their recognition of voices, but what if there is loud background music? What if there are a lot of conversations going on? Distinguishing one person's voice from another may be impossible. There are options to overcome these challenges. A blind person could announce out-loud who they are looking for and depend on the helpfulness of others to assist them, but what if the person they want to approach is a world-renowned professor or a romantic interest? It is less than appealing, and sometimes embarrassing, to have to announce this to everyone. Often, the easiest course of action for a blind person is to wait in one place for someone to approach you. However, the visually disabled are people too. They want to have the option to be independent and they appreciate their privacy just like anyone else.

Facial recognition is a field of research in computer science devoted to programming computers to locate and identify faces from pictures, pre-recorded videos, and even live video feeds. This is not easy with the possible variations in lighting, quality of picture, and directions people face, but great strides have been made towards performing this task with accurate results. Its greatest application has been towards security in high traffic places like the airport. Facial recognition can assist the Transportation Security Administration (TSA) in identifying terrorists and criminals. The power of these recent advancements in facial recognition can be harvested into a visual social aid for the blind.

Socialeyes [ND10] is a social tool for the visually disabled that is currently being designed and developed at Colorado State University. It will ultimately utilize their state-of-the-art facial recognition program, FaceL [BBD09], to assist users in recognizing and locating specific individuals in a small room setting of approximately ten occupants. In its final form, the tool will be used through a mobile device such as a cell phone. In addition to creating a useful tool, the developers are concerned with creating a convenient and easy to learn interface for its users.

The visually disabled process information and feedback with very little or no sight. Some forms of feedback are more useful than others. In fact, sighted people often take for granted the ease in which they can handle daily navigation events such as avoiding a crack in the sidewalk or maneuvering in a room towards peers. To help with these situations, many tools have been developed to help the visually disabled travel such as the cane or global positioning systems, but what makes these tools useful and easy to use? What qualities make it a burden to use? Designing tools requires an understanding of what the user needs, where the user will use the tool, and how the user processes information.

We begin a case study, with the assistance of Adam Campfield, to explore and discuss the attributes that make a tool for the visually disabled useful. Adam Campfield is a blind, undergraduate Applied Computing Technology major at Colorado State University. His gregarious personality and interest in Human Computer Interaction made him an ideal volunteer for this project. This pilot case study explores two scenarios. The first scenario tasked Adam with finding any person in a given room and navigating to them. The results were used to develop a reasonable Socialeyes Emulator. The second scenario involved locating and navigating to a specific person. Its success was based on the time to complete the task.

While the information gathered here is based on the feedback and observations from a single blind user, it is more applicable and useful than feedback gathered from several blind-folded, sighted participants. There are distinct differences between a sighted person who is blind-folded and a blind person who has learned to adapt to the lack of visual cues. These differences include the way they move and the way they interpret information. As the Socialeyes system is intended for users who are visually disabled, it follows that visually disabled users would provide the most relevant feedback.

This study shows that Socialeyes should complement the natural person finding techniques that the visually disabled have developed. The type, amount, and form of feedback provided to the user are all important. Feedback needs to be concise, prompt, contain useful information, provided in manageable amounts, and in at least two forms. The two forms of feedback used in these trials were haptic and audible. From the results of this study, future work can progress to using lighter equipment, conducting objective studies of the factors of a usable interface for the visually disabled, and researching the relationship between models of human behavior and human computer interaction.

The rest of this thesis is organized as follows. The following chapter presents past work on travel aids, object and facial recognition, and characteristics of a usable tool for the visually disabled. Next the usability study and its results will be presented. Lastly we'll present the conclusions and ideas for future research.

Chapter 2

Literature Review

There exist many tools to assist the visually disabled in traveling. The white cane and guide dog are the classic options. Each provides the user with information on obstacles, changes within the path, and lets drivers know that the person is blind so caution is advised. In recent years global positioning systems have been integrated with mobile devices where synthetic speech is used to convey information useful for travel [May00]. The newest innovations include providing information on public transportation systems, their schedules, and reports of possible delays.

Goto and Kambayashi [GK02] developed a travel tool that utilizes a mobile database system and provides passengers with the ability to search routes and retrieve fare information, station maps, operation schedules, and information on vehicle facilities. Their system was designed with the following key concepts in mind; the dynamic nature of public transportation systems, that all users are different, and that all users have different needs. Buses, trains, and even subways can run behind schedule or break down. This is the kind of information that needs to be conveyed to passengers and the way it is conveyed is through a personalized user interface. For the visually disabled, a slight variation on the original system was proposed. In order to help the visually disabled navigate public transportation stations, radio-frequency identification (RFID) data tags could be embedded into the floor. Each tag would provide information on its location that could be read by a cane and transmitted to a mobile device.

Other tools being developed for travel utilize the research done in computer vision. Computer vision studies how machines can take images or video and analyze the contents. Often it is used to teach machines how to recognize objects and to learn more about how humans process what they see. Silapachote et. al [SWH+05] began work on a system that could find and recognize street signs. It utilizes a database of pictures of different signs and is designed to work on a mobile device. As of 2005 there has not been any work done to make this usable for the visually disabled, but there is potential for it.

Coughlan, Manduchi, and Shen [CMS06] worked on a system where visually disabled users could take an off-the-shelf cellular phone and use computer vision to navigate within a building. Specifically, computer vision was used to detect colored signs and to read their associated barcodes. These signs were robust under changes in lighting and were used to label different locations within the building. Tests were performed where the subject used the system to locate the colored signs. In order to prevent the user from using any previous knowledge of the building, the location of the signs did not necessarily correspond with the location it described. The system communicated with the subject through pre-recorded audio files identifying the name of the location found. Overall, it was a successful series of tests with the potential for real-world applications.

Instead of object recognition, some tools are designed to use facial recognition. This involves the system detecting faces and then identifying them. FaceL [BBD09] is a state of the art, real time facial recognition and labeling program developed at Colorado State University. It can track multiple faces accurately over a live video stream. Up until recently, facial recognition systems needed a substantial amount of time to process images or saved videos. With FaceL, the quick processing allows users to receive information almost instantly about who is in the video. The interest here is in using this advanced facial recognition system to identify people by their faces for users who have difficulty or cannot do this on their own.

Although many everyday tools are being developed for the visually disabled, there has not been a large amount of research done on usability. The features and forms of feedback for visual aids need to be investigated. A possible feature to include is personalized settings such as what voice to use for text to speech programs. Users could also set what speed information is read to them. Perhaps having messages read out loud is not the preferred form of feedback. Some users could prefer a more discrete form such as vibrations or audible tones. In addition to features and forms, the amounts and timing of feedback that make a tool accessible and usable need to be thoroughly analyzed. The few studies done involved testing the different types of feedback or discovering what characteristics of verbal communication most benefit the visually disabled.

In 2000, Ross & Blasch [RB00] evaluated the effectiveness of orientation interfaces in wearable computers. Their results showed that the combination of tactile and speech feedback was most effective in communicating with visually disabled users. Crowder & Morton [CM69] suggested that speech sounds take up more space in short term memory than non-speech sounds. They also thought that non-speech sounds disrupt the cognitive processes less than speech. This suggests that using non-speech can be an effective substitute for some speech feedback.

Pitt & Edwards [IA96] analyzed speech synthesizers combined with screen readers and determined several interesting things. Speech based interfaces lack cues that human speech naturally includes such as variations in pitch, intonation, volume, and the appropriate incorporation of pauses. For example, many languages utilize a change in the pitch of their voice to indicate that a question is being asked. Increased volume can indicate urgency or fear. Well placed pauses can put emphasis on words or phrases while inappropriately placed pauses can disrupt reaction time by multiple seconds.

In Pitt & Edwards' experiment, a version of hangman was used that both sighted and unsighted users could play. A screen reader with voice synthesizer was provided and the monitor was hidden from the players. In general, all participants thought there was too much speech being provided as feedback and suggested the use of non-speech sounds as substitutes. Too much speech was distracting the players from being productive. Another observation was made that ”hot keys,” buttons to push when specific feedback is desired, would make the interface easier to use.

A study was done in 2005 on the different characteristics of verbal directions that were most useful to sighted people versus visually disabled people. Bradley & Dunlop [BD05] took sixteen volunteers, eight sighted and eight visually disabled, and had them travel to four different locations within a city. Two of the locations could be found using verbal directions generated from a sighted person’s descriptions while the other two locations could be found using verbal directions from a visually disabled person’s description. Each set of directions was recorded on a MiniDisk and all volunteers tried both sets of directions. In general the visually disabled volunteers reached the landmarks quicker when given directions from a visually disabled person’s descriptions, although they took longer to reach all of the landmarks than the sighted volunteers regardless of the set of directions. It was noted that the visually disabled used more structural information, like road or monument, and descriptive information such as steep or tall. The sighted volunteers used more textual-structural and street information, like the names of nearby shops or cross-streets, providing more evidence that further studies need to be conducted on what information is most useful for blind users.

In Canada, Strothotte et. al began developing the MoBIC Travel Aid (MoTa) to help the blind and elderly with planning and executing independent travel plans [SFM^{+}96]. MoTa was made of two parts. The first part was called the MoBIC Pre-Journey System (MoPS) and was designed to look up digital maps, public transportation info, travel times, and more. This information could be transferred to the second part of the system called the MoBIC Outdoor System (MoODS). MoODS was designed to complement the common travel tools such as a cane or seeing eye dog. It uses GPS and a mobile telecommunication facility to notify users when they are deviating from the path and help them find their way back to the path. When design began on this complicated system, potential users were interviewed and asked about preferences and needs with respect to the proposed aid. The interviewees wanted to not only be able to control what and how much feedback was communicated, but how much detail was in the feedback. They also discouraged the developers from having their users wear headphones since it could distract from environmental sounds which usually aid in navigation. Instead a separate mobile piece could be provided that users could hold up to their ear when feedback is desired.

Initial testing for the MoBIC Travel Aid system included five tutorials training users on how to explore areas, plan a journey, and make the journey. Users were hopeful and made the observation that while exact distances provided specific information, it wasn't always helpful for users who were unable to judge distances accurately.

Based on the research done, in order to create a usable facial recognition system to help the visually disabled navigate, accurate and helpful feedback needs to be provided. Speech feedback needs to be paired with another form of feedback such as audible sounds or tactile feedback. The following chapter presents our exploration of human computer interaction for the visually disabled beginning with the design of feedback messages.

Chapter 3

Methodology

In order to determine the most useful forms of feedback for users of the Socialeyes system, our case study consists of trials conducted within two scenarios. The first scenario was the simplest while the second scenario built upon it. The first scenario involved the user attempting to recognize and locate any individual in a room while the second scenario tasked the user with recognizing and locating a specific individual. This design allowed the most information gain as it relates to the Socialeyes system.

The room used for testing is a small discussion style classroom. Its quiet location received very little noise from outside sources. The room contained desks, that were sometimes used as obstacles, and chairs for the volunteers to sit in. Each trial explicitly states whether obstacles were used.

The system consisted of a webcamera, earphones, and two laptops. One laptop was used by the tester and the other was carried by Adam Campfield, an undergraduate assisting in all of the trials. Adam carried his cane in hand and the laptop in a backpack. A webcamera was used in every trial, but the placement of it varied. The earphones were plugged into the laptop to ensure discretion and the privacy of the messages sent. Over the course of this study, the system evolved and took on new pieces of equipment. Figure. 3.1 shows the current version of the system in use during a trial. Section 3.1 will describe the components of this system in more detail.

Volunteers were instructed to find a chair and move themselves to an arbitrary loca-

Figure 3.1: Current version of the Socialeyes Emulator System; two laptops, earphones, webcamera attached to a headband, cane, and Wiimote.

Trial	Attempts	Successes
1	1	1
2	2	1
3	2	2
4	3	0
All	8	4

Table 3.1: Scenario One Summary: Success Count

tion in the room. They were asked not to move or make noise, which included turning off cell phones, with the exception that they could talk if they felt someone was in danger of injury. They were also informed that Adam would be moving around the room with the goal of trying to locate them.

The forms of feedback used varied slightly between trials. In general, the user needed to be notified of three major events; when a face was found, recognized, or lost from the view of the camera. Supplemental information such as distance measurements and relative directions aided in the subsequent user action.

3.1 Scenario One

The simplest task for a user of Socialeyes is to find a single person. To maintain this simplicity, the user did not need to find a specific person, but instead any person in the room was satisfactory. We evaluated success based on whether Adam was able to complete this task. Out of the four trials, Adam was able to successfully find a person in four out of the eight attempts. The data is summarized in Table. 3.1.

Trial One

In the first trial of Scenario One, the emulator of Socialeyes used three main messages. The first two messages were Face Detected and Face Lost. Both messages were followed by a relative direction. Eight directions were used and were defined as in Figure. 3.2.

Up Left	Up	Up Right
Left		Right
Down Left	Down	Down Left

Figure 3.2: Directions used in Scenario One: Trial One

If a face was detected in the upper left box of the screen, a message would be sent letting the user know that a face was detected in the upper left of the webcamera's view. Similarly if a face was lost from the right side of the camera's view, a message would be sent indicating that a face was lost to the right. The third main message was Face Recognized. This message was accompanied by an estimated distance to the person recognized. Each message was read using a voice synthesizer. No obstacles were used in this trial.

Due to the unavailability of a conference room, the initial testing was conducted in a small lounge area. Two volunteers participated in this trial and situated themselves in different locations in the lounge. As the developers envision the final product to be contained within a cell phone, it made sense to have Adam carry the webcamera in his hands.

Once the attempt began the webcamera sent video and sound to the tester's laptop over Google's Voice and Video Chat program [Goo11]. This program was chosen because of its video chatting abilities and low price. The tester responded to the transmitted video by composing a message using the previously defined feedback. Each message was sent via a socket program and was created using a keyboard input oriented interface which first selected the type of message and then the corresponding arguments. The types of message included Face Detected, Face Lost, and Face Recognized. In the case of the first two types, the corresponding argument was a direction. For Face Recognized the appropriate feedback was the name of the individual recognized in the video and an approximation of how far away the individual was. The feedback was then transmitted to Adam's laptop and read to him by a voice synthesizer. Adam was successful in locating one of the volunteers and several key observations were made on how to improve the interface.

Notably, the messages read by the voice synthesizer were clear and understandable. However, Adam observed that the messages itself were long. By the time the message was fully read out, Adam found himself pointing the camera at a different scene than the message was intended for. The most significant flaw in the system was the amount of lag between a face being detected, lost or recognized and a message received by Adam. Part of this was due to the strength of the connection to Colorado State University's wireless network and the other part was due to the response time between the tester seeing an event on the video feed and responding appropriately to it. The video feed often froze during this trial and the feedback messages took several seconds to transmit. Additionally, each laptop ran significantly slower as it seemed to have difficulty handling both the video chat program and the separate socket program used to transmit feedback.

The tester noted that the system was missing a necessary message to indicate that a face was found in the center of the camera's view. Without this message the tester was unable to inform Adam he was on the correct path to finding a person. It was also found that the keyboard input oriented interface meant the tester was spending a significant amount of time trying to construct the messages to send.

Trial Two

From Trial One, we learned that the initial design of the Socialeyes Emulator had significant weaknesses including the lag between an event being detected and the user receiving feedback about it. In order to minimize this lag, Trial Two does not use the socket program to send messages. Instead, messages are sent over the video chatting program. Unfortunately, Google's Voice and Video Chat program was not accessible. This meant that the commonly used screen reader program JAWS [FS11] could not read from it. Therefore the video chatting program used in Trial Two is Skype [Sky11]. A consequence of this change was that the feedback messages were not set up in advance and were not standardized, but the tester could send more flexible messages and possibly communicate more to the user.

This trial also attempts to answer the question, "Does a range of distances work as feedback or are discrete values more useful?"

The second trial consisted of two tests. Obstacles were not used in either test. The system still consisted of two laptops and Adam carried his cane and a webcamera. Most of the feedback messages were kept the same, but some were simplified including the Face Recognized message. The new message did not announce that a face was recognized, but rather just communicated whose face it was and an estimate of how far away that person was from the user. A face detected message was also added to indicate that a face was found in the center of the webcamera's view.

The first test consisted of three volunteers. Once the test began, feedback was sent to Adam through Skype and JAWS was used to read out loud the information. Unfortunately Adam did not receive any of these feedback messages due to the placement of the microphone. There was also no built in feature to allow Adam to communicate this problem to the system. This event led to the conclusion that not only does the system need to be able to communicate with the user, but the user needs to be able to provide the system with information such as failure or incorrect output.

The second test had two volunteers. During the course of this run, the face in the view of the webcamera was lost from the bottom of the screen five times. This was most likely caused by the lag between feedback being sent and received. In the end, Adam was able to find a person successfully. When a volunteer was recognized by the emulator, the system provided an estimate of the distance to the volunteer. From this, Adam observed that specific, discrete values indicating distance to a person is more helpful than a range of distances. Unfortunately distance is a difficult thing to estimate accurately and not all users will be as adept at gauging distances as Adam is. Adam also reported that the user needs to be able to communicate to the system to stop looking for a person it recognizes.

Adam observed that while the feedback messages being read to him were useful, it was creating a lot of noise. As this was the only form of feedback the user was receiving, it required focused attention for an extended period of time. This meant that noises from the room and the potential noises from the volunteers that would normally aid in finding their locations were actually distractions from the feedback.

We also found that Adam moved his head around a significant amount as he tried to find a person. Part of this could have been attributed to the fact that he would tilt his head to hear sounds better. It was also possible that since Adam doesn't use sight, the movements of his head are independent of his attempts to move in a specific direction.

Trial Three

As was identified in the previous trial, having all of the feedback communicated to the user in an audible form was overwhelming. This effect, called audio spamming, required a great deal of concentration from the user and tended to make the task of locating a person more difficult.

Trial Three aimed to cut down on the audio spam, but still communicate the same amount of information as before, by utilizing haptic feedback via Wii Remote Control.

WiiRemoteJ is a Java library developed by Michael Diamond [Dia09]. Its purpose is to provide a java implementation to access the features of a Wii Remote Control (Wiimote). It can make a Wii Remote vibrate, play sounds, and connect the buttons to different actions. WiiRemoteJ is used in conjunction with an implementation of JSR082, a Java Bluetooth API, such as BlueCove [Blu10].

In order to have a hand free to hold the Wiimote, the webcamera was attached to a headband for Adam to wear. This also provided the system with a view of what Adam's head was directed towards. In addition, the feedback was changed such that if the Wiimote did not vibrate, it meant that there was no face in the view of the webcamera while vibrations indicated there was someone there.

In the first test the Wiimote used modulated vibrations to indicate the approximate distance from the user to the target. The faster the vibrations, the closer the user was to the target. In this test there were two volunteers and no obstacles. Once the test began, Adam observed that the modulated vibrations were inconsistent. The vibrations were not evenly spaced and the length of the vibrations varied. There was no distinct difference from one speed to another. By adapting and using the Wiimote in a binary fashion, with pulses indicating face found and no pulse indicating no face in view, Adam was still able to locate a person successfully.

In the second test the Wiimote's behavior was modified to use a different number of vibration pulses to indicate distance from the user to the target. The more pulses, the closer the user was to the target. One pulse indicated a face was found. Two pulses indicated that a face was recognized, but the user was approximately ten to fifteen feet away. Three pulses indicated the user was approximately five to ten feet away from the target and four pulses indicated a distance of one to five feet. Adam observed that the modified Wiimote behavior and the balance between tactile feedback and audible feedback was much easier to handle. He preferred it over an entirely audio based feedback system.

In test two, there was only one volunteer and a single table used as an obstacle between the user and the target. Adam was able to successfully find the volunteer and observed that the different number of vibrations was better feedback than the inconsistent modulated vibrations. The obstacle, however, brought about a new problem. In order to navigate around the table, Adam turned his body and head and lost the volunteer's face from the view of the camera.

According to Adam, when a blind person is searching for something or someone and becomes disoriented, they aim for the general direction they believe the target is in. For sighted people reacquiring a target is easy, but for the blind this can be quite a difficult task. The system at this point did not have feedback as to whether Adam should turn left or right to reacquire the target. In fact, the only way that Adam knew he had lost the target was that the Wiimote stopped vibrating.

It was also observed in this run that Adam moved his head significantly less than in the past trials. This could be attributed to the quicker response in feedback.

Trial Four

Trial Four had three tests. The Wiimote was still utilized along with the different number of vibration pulses to indicate distance from the user to the target, but the system utilized two new features. The first was TeamViewer [Gmb11], a free PC program designed to create a remote desktop connection between computers without the need for administrative passwords. This program allowed the system to be centralized on one laptop and the trials to run more consistently with less troubleshooting. It also eliminated the need for an accessible video chat program as the webcamera could be accessed directly. The second new feature was pre-recorded audio files of a male voice saying the words left and right. One of these files was played when a face was lost from the webcamera's view. The relative direction, left and right, indicated which direction the user should turn to reacquire a target.

Test one had one volunteer and one obstacle. While Adam was on the right track to locating the target, he forgot what the different number of pulses indicated. This led to him being approximately ten feet away from the target when he thought he was about two feet away.

Test two had two volunteers. While navigating about the room, the target's face was lost to the left of the webcamera's view. Adam was in the process of turning left when the system notified him to turn left to recover the lost face. His assumption was that there was a delay and so began to turn to the right to recapture the face. Unfortunately, his anticipation of the delay led to him not finding the original target and colliding with the other volunteer. Obviously timely feedback is vital for this trial, but feedback for user error also needs to be incorporated into the system.

Test three also had two volunteers and ended with similar results. In anticipation of the delay between the system detecting a face in the camera's view and the receipt of feedback, Adam turned the opposite direction from what he was instructed to do. This resulted in him circling around one of the volunteers, but never recapturing the volunteer's face in the camera's view.

We found that the fewer number of things the user had to remember, the easier the system was to use. In the next trial, the number of settings indicating distance to a target was minimized, pre-recorded audio files were played to remind the user of different settings, and the balance between haptic feedback and audio feedback was maintained.

Trial	Test	Time to Locate Face	Time to Locate Target	Total Time
1	1	2:32	0:10	2:46
1	2	1:05	0:52	1:57
2	1	0:46	0:14	1:00
2	2	0:42	0:16	0:58
2	3	0:59	0:11	1:10
3	1	2:14	0:10	2:24
3	2	1:13	0:05	1:18

Table 3.2: Scenario Two Summary: Times (min:seconds) to Find a Specific Person

3.2 Scenario Two

The next trial moves on to Scenario Two where we introduce a multiple person setting. The user was tasked with finding a specific person in a room of two people. In the previous scenario, no testing was done to see if a target was more difficult to find if they were sitting or standing. Scenario Two was an ideal time to test this since there were multiple people in the room and a specific target could be specified. The difficulty of the task was evaluated based on the amount of time it took to complete it. See Table. 3.2 for a summary of the results.

As in Section. 3.1, the system was composed of two laptops, a cane, earphones, a webcamera, Wiimote, and headband. Team Viewer was used along with Logitech's webcamera software to establish a connection to the webcamera. To make the system more usable, pre-recorded audio files were played to the user the first time each event occurred. Each audio clip provided a reminder to the user of what the different forms of feedback meant.

Adam is very adept at being able to find people he knows as he can recognize their voices and the way they move. To prevent Adam from inadvertently using any of his previous knowledge of the volunteers, code names were assigned. One volunteer was designated Alpha and the other Bravo. Adam was not informed which volunteer would be sitting and which one would be standing.

Each volunteer was told their code name and asked to choose a location in the room. One volunteer was asked to stand while the other was asked to sit. As in Scenario One, the volunteers were asked to make noise and move around as little as possible.

Trial One

There were two tests in Trial One. At the beginning of each test, Adam was informed which target he was looking for. The Wiimote was still used to provide haptic feedback on the distance from Adam to the target. During the trial, the three main messages were provided as feedback, but were slightly modified. When a face was lost or recognized, the system reported whether it was Bravo's face or Alpha's by playing a pre-recorded audio file of a male voice saying the appropriate code name.

The target of test one was Bravo who was sitting. It took Adam approximately two minutes and forty-six seconds to identify and locate Bravo. The most difficult part of this task was when there was no face in the view of the camera. The Wiimote provided no vibration, since there was no face, and the system could not provide information on which way the user should turn to find someone. Approximately two minutes and thirty-two seconds of the total time were spent trying to locate any face. Once the intended face was found, it took very little time to travel to the target.

In the second test, the target was Alpha who was standing. This test took a total of one minute and fifty-seven seconds, thirty-five seconds of which was spent looking for any face. The first face found was Bravo's and Alpha's face was not found until one minute and five seconds into task. Notably it took significantly less time to find the standing target, Alpha, than to find the sitting target, Bravo.

With sighted people, peripheral vision is a tremendous asset that allows us to see more than what is in front of us. A webcamera does not have peripheral vision and so the system is unable to provide any clues as to where to look for a possible face. In order to compensate for this, the following trial utilizes audio clips to identify where people are in the room.

Trial Two

When a blind person is searching for someone in a room, the noises made by conversations can be helpful in pinpointing their location. In the past trials, the volunteers were instructed to make as little noise as possible to ensure that Adam was not using any previous knowledge of the volunteers to locate them. However, this made the task of locating a person very difficult as the system was unable to give instructions on where to begin looking.

In order to provide information on the location of the volunteers, but not their identities, each volunteer was given a mobile device that could play a short audio file. The volunteers were instructed to begin playing the audio file, on repeat, at the beginning of each of the three tests. This provided Adam information on where people were in the room, but not who the people were.

In test one the target, Alpha, was sitting. Even though he first found Bravo, it took Adam approximately one minute to locate and correctly identify Alpha. This was a significant time decrease from Trial One as can be seen in Table. 3.2.

In test two, the target was Bravo who was standing. The test took a total of fifty-eight seconds. Approximately fourty-two seconds of the total time was spent looking for any face. Again, these were much shorter times than in the previous trial.

At this point it was probable that Adam may have associated each audio clip with the code name. Without informing him, the volunteers traded media players. The goal for test three was, again, to find Bravo. The test took approximately one minute and ten seconds to complete.

The use of audio clips made locating the volunteers a much easier task. The next step was to have the volunteers face in different directions away from Adam.

Trial Three

Each volunteer in Trial Three was given the same instructions as in previous trials except this time they were asked to face in different directions that were not towards Adam. FaceL is unable to detect and identify people from the back of their heads so the idea here was to determine how difficult it would be to navigate around a person to find a face.

There were two tests and two volunteers per test. In the first test Bravo, the target, was standing and facing to the left. Alpha was sitting and facing to the right. At first Adam approached Alpha and needed to circle around to find Alpha's face. This took some time as Adam had difficulty getting Alpha's face in the view of the webcamera. This may have been due to the difference in height, since Alpha was sitting while Adam was standing, and could also be attributed to how close Adam was to Alpha. Alpha's face filled the webcamera screen to an unrecognizable point. It took Adam approximately two minutes and fourteen seconds to find the face of Bravo. From that point, it took ten seconds to identify that person as Bravo.

In the second test both volunteers were standing and facing new directions. Alpha was the target. It took Adam approximately one minute and eighteen seconds to find Alpha.

Adam observed from this trial that the difficulty increases when searching for people whose faces are not all at the same height. He suggested that we add to the Face Lost feedback message a way to determine if the face was lost from the top or bottom of the webcamera's view. This will be discussed further in Section. 5.

The times from Trial Three were significantly better than the times from Trial One (See Table. 3.2). Having noises to help narrow down the search area of where people are can greatly ease the task of locating a specific person. Since real-world situations will have noises, this will be a benefit to the users of our system.

Chapter 4

Conclusion

Adam's opinion of a visual social aid was one of excitement. He commented that he had found himself searching for people at parties and a tool like Socialeyes would be helpful. After participating in this pilot study, he remained enthusiastic. Adam thought we were making important progress on designing the features of the interface, but was eager to move away from the laptops and head bound webcamera and move on to lighter and more sophisticated pieces of equipment like a mobile device.

Several important concepts have resulted from this case study including how, how much, and what kind of feedback needs to be conveyed. Our study shows that the response time of feedback is vital. Once the system detects an event such as Face Lost, it needs to notify the user in a prompt manner. The consequence of not doing so, as we've discovered in our initial testing described in Section. 3.1 is that the user will not be able to accurately associate the feedback with the event. This leads to a decrease in usability as well as possible failure to find a person in the room.

Another finding is that feedback from the system needs to be concise. There is a lot of information that needs to be communicated to the user, but verbose messages can distract from the intended task. The more time a user spends listening and interpreting a message means less time spent listening to other things like a friend trying to start a conversation with them. Attempting to listen to both the feedback from the system and from the natural sounds of the room can mean less focus on the task at hand. Adam observed in Scenario One's first trial, described in Section. 3.1, that the messages were too long and that he had to consciously focus in order to hear the entire message. The result was a halt in movement while receiving feedback and meant that the feedback was serving as a distraction from the environmental noises that Adam normally used when looking for someone in the room.

This study also suggests that the content of the feedback provided is just as important as how it is communicated. In Section. 3.1, Scenario One's second trial, we found that Adam preferred discrete values to indicate distance to the target rather than using a range of values. Other studies, such as Strothotte et. al [SFM+96], agree with Adam's opinion, but caution that not all users are alike in being able to accurately judge distances. Perhaps the option of distance feedback as a range of values versus discrete values should be left as personalization of the user interface.

Another result found relates to how much feedback is provided. When the system was programmed to provide feedback only in an audible format, Adam found the audio spam distracting from his task of finding a person. The results of Trial Three in Scenario One suggests that haptic feedback can be substituted for some of the audio feedback. More specifically, haptic feedback could be used to indicate that a face was detected and how far away that face was. This reduced the amount of audio feedback while still providing the necessary information to help Adam with his task.

In fact, having too much feedback of any kind can be distracting and sometimes overwhelming. The visually disabled have adapted to not having sight and in the process learned many techniques to navigate without the use of technology. A visual social aid does not need to provide information on everything occurring in the webcamera's view as the users are capable of gathering some information on their own. In Section. 3.2 we found that even though the webcamera used in testing lacked the peripheral vision needed to provide Adam with directions on where to look for faces, noises from the people in the room could provide sound based localization. This information could be used to narrow down the search space immensely.

Overall this study has shown that Socialeyes should complement the natural person finding techniques that visually disabled users already have. When visually disabled people travel, whether it be within a building or across town, the noises generated by people and other objects help to find their destination. A person could narrow down the search space within a small room quite quickly by using the surrounding noise. The vast time differences between Trial One and Trial Two as shown in Table. 3.2, attest to that. Socialeyes should assist users after that point to identify the people within sight.

Chapter 5

Future Work

This pilot study presents a starting place for future research and advances. For example there are many scenarios that have not been tested yet including having more than two volunteers within a trial or having people in the room that the system does not recognize. There are also many challenges yet to face. Incorporating a feedback message into the system that informs the user a face was lost from the top or bottom of the screen could shorten the amount of time to recapture a target's face. The additional information could also assist in finding a face while navigating around obstacles. Users may prefer that the system be more discrete. Other people should not be able to tell who the user is looking for or what kind of feedback the system is providing. Initially the webcamera was held in Adam's hands and then was mounted on Adam's head via a headband, but this is not ideal as the webcamera is noticeable and can draw unwanted attention to the user. Future trials could test how moving the camera to the torso affects the usability of the system.

Along with testing new scenarios there is progress to be made towards using better equipment. The current version of the emulator requires the user to carry a laptop on their back, Wiimote in one hand, cane in the other, and wear a webcamera attached to the forehead by a headband. Adam's comments on the overall study was that it had potential, but he was looking forward to using equipment that was easier to carry. Instead of the laptop and Wiimote, a mobile device such as a cellular phone could be used. The cellular phone has its own computing power as well as the ability to vibrate for different events. Additionally, most cell phones feature a camera that could be used in place of the webcamera. This raises another usability issue, though, as the user would have to hold the mobile device out in front of the torso and scan the room with it. This is unnatural and strange to other people in the room. Instead of using the cellular phone's camera, a pair of sunglasses could be used with a built in camera. This would draw less attention to the user and would be more comfortable to carry.

The study presented has been significantly subjective. The results are based on one user's opinion, but we have identified important factors that need to be further researched in order to develop a visual social aid. A systematic, objective study needs to be conducted over a larger blind population on factors such as how much delay between the occurrence of an event and the user receiving feedback about it the user can tolerate. What is the largest height difference the current system can handle? How much error from the facial recognition program can the user tolerate? Throughout this study there has been an assumption that the facial recognition program to be integrated into the Socialeyes system will perform perfectly. That will not be the case and it's important to know how much this effects usability.

As Socialeyes is intended for situations involving social interaction, the study of human behavior could be significant. Human behavior has been studied and analyzed for years. The resulting models describe different personality types that provide insight into the way people interact with each other. Our personality influences how we communicate and handle conflicts and therefore on how we socialize with friends, co-workers, and family. A natural extension is to see how the differences in human behavior affect human computer interaction (HCI).

A useable visual social aid, like Socialeyes, has the potential to make finding people an easier task for the visually disabled. The blind could take it upon themselves to find someone they know at a party instead of waiting for someone to approach them. A professor could use this tool to remember the names of the hundreds of students in her course. The elderly could use it to test and strengthen their memory. Further research is needed to explore the different applications Socialeyes could be used in and how to make its interface intuitive.

REFERENCES

[BBD09] D. Bolme, R. Beveridge, and B. Draper. Facel: Facile face labeling. In *International Conference on Vision Systems*, pages 21–32, October 2009.

[BD05] Nicholas A. Bradley and Mark D. Dunlop. An experimental investigation into wayfinding directions for visually impaired people. *Comp*, 9:9–395, 2005.

[Blu10] BlueCove. Bluecove. `http://bluecove.sourceforge.net/`, 2010.

[CM69] R.G. Crowder and J. Morton. Precategorical acoustic storage (pas). *Perception & Psychophysics*, 5:365–373, 1969.

[CMS06] James Coughlan, Roberto M, and Huiying Shen. Cell phone-based wayfinding for the visually impaired. In *st Int. Workshop on Mobile Vision*, 2006.

[Dia09] Michael Diamond. Wiiremotej. 2009.

[FS11] Inc. Freedom Scientific. Jaws for window screen reading software. `http://www.freedomscientific.com/products/fs/jaws-product-page.asp`, 2011.

[GK02] Koichi Goto and Yahiko Kambayashi. A new passenger support system for public transport using mobile database access. pages 908–919, August 2002.

[Gmb11] TeamViewer GmbH. Teamviewer. 2011.

[Goo11] Google. Google chat help. `http://www.google.com/support/chat/bin/answer.py?hl=en&answer=159499`, 2011.

[IA96] Pitt IJ and Edwards ADN. Improving the usability of speech-based interfaces for blind users. In *2nd annual ACM conference on assistive technologies*, pages 124–130, Vancouver, Canada, 1996.

[May00] Michael May. Accessible gps navigation and digital map information for blind consumers. In *Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000)*, pages 317–319, Salt Lake City, UT, September 2000.

[ND10] Erin Nagoshi and Bruce Draper. Socialeyes. 2010.

[RB00] David A. Ross and Bruce B. Blasch. Evaluation of orientation interfaces for wearable computers. In *Proceedings of the 4th IEEE International Symposium on Wearable Computers*, ISWC '00, pages 51–, Washington, DC, USA, 2000. IEEE Computer Society.

[SFM+96] Thomas Strothotte, Steffi Fritz, Rainer Michel, Andreas Raab, Helen Petrie, Valerie Johnson, Lars Reichert, and Axel Schalt. Development of dialogue systems for a mobility aid for blind people: initial design and usability testing. In *Proceedings of the second annual ACM conference on Assistive technologies*, Assets '96, pages 139–144, New York, NY, USA, 1996. ACM.

[Sky11] Skype. Skype. `http://www.skype.com/`, 2011.

[SWH+05] P. Silapachote, J. Weinman, A. Hanson, M.A. Mattar, and R. Weiss. Automatic sign detection and recognition in natural scenes. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, page 27, 2005.