ABSTRACT

Title of Thesis: FACIAL AND EXPRESSION RECOGNITION FOR THE BLIND USING COMPUTER VISION

Authors: Douglas Astler, Harrison Chau, Kailin Hsu, Alvin Hua, Andrew Kannan, Lydia Lei, Melissa Nathanson, Esmaeel Paryavi, Michelle Rosen, Hayato Unno, Carol Wang, Khadija Zaidi, and Xuemin Zhang.

Directed by: Professors Rama Chellappa* and Cha-Min Tang**
*Department of Electrical and Computer Engineering
*UMD Institute for Advanced Computer Studies (UMIACS)
**University of Maryland School of Medicine

The majority of communication between humans is comprised of nonverbal cues. Visual cues, such as expressions and nodding, are not readily accessible to the blind. Our team has developed an assistive device based on computer vision, which relays facial recognition and expression information to a blind user and is designed to store images of people the user frequently interacts with and analyze faces for expressions. The control a user has over his/her surroundings while receiving real-time feedback contributes to a unique device. In order to design a device suited best to the blind user's needs, we engaged sighted and blind participants in surveys and interviews to understand their views and preferences regarding the availability of a computer vision system that, in real-time, can provide information about identity and expressions of humans. This thesis discusses the development methodology, the selection of algorithms for recognizing faces and expressions and physical designs of the device, and the results on subject tests with blind participants to gauge the effectiveness of our design.

FACIAL AND EXPRESSION RECOGNITION FOR THE BLIND USING COMPUTER VISION

by

Douglas Astler, Harrison Chau, Kailin Hsu, Alvin Hua, Andrew Kannan, Lydia Lei, Melissa Nathanson, Esmaeel Paryavi, Michelle Rosen, Hayato Unno, Carol Wang, Khadija Zaidi, and Xuemin Zhang.

Thesis submitted in partial fulfillment of the requirements of the
Gemstone Program
University of Maryland, College Park
2012

Advisory Committee:

Doctors Rama Chellappa and Cha-Min Tang, Mentors

Dr. Yaser Yacoob
Dr. Larry Davis
Dr. Ankur Srivastava
Dr. P. Jonathon Phillips
Ms. Sima Teheri
Ms. Anne Taylor

ACKNOWLEDGEMENTS

In our three years of research on developing an assistive device for the blind, we were continuously humbled by the gracious support of all those around us. We are indebted to Anne Taylor and others from the National Federation of the Blind for their advice and encouragement. We thank the participants in our research who bestowed their time and experience in answering our surveys, interviews and in trying our device. We are grateful to our panel of discussants and advisors who patiently answered our questions and shared their knowledge with us. Without them we could not have created this device. Mostly, we sincerely thank our mentors, Dr. Rama Chellappa and Dr. Cha-Min Tang for their guidance and investment in us during our three years as Team FACE. We thank Dr. Wallace, Dr. Thomas, and the entire Gemstone Honors Program for their continued support during our four years as undergraduates. Finally, we thank our parents and families for understanding our ambitions and believing in us in our 21+ years as their children.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	Application Programming Interface
CV	Computer Vision
DLL	Dynamic Link Library
GUI	Graphical User Interface
IRB	Institutional Review Board
LBP	Linear Binary Pattern
NFB	National Federation for the Blind
PCA	Principal Component Analysis
SDK	Software Development Kit

Chapter 1

Introduction

A face provides genuine insight into expressions and emotions. The distances between facial features and rapid variations provide more reliable cues than verbal communication. Thus, human communication can be divided into two parts: verbal and nonverbal. Here, nonverbal communication is described as the process of communicating through facial expressions, eye contact, and nodding in addition to nonverbal aspects of speech such as tone and volume. Over 80% of all communication is carried out through these nonverbal messages [1].

For the visually impaired, a face provides little to no insight into a person. According to the World Health Organization, there are about 284 million people in the world who are visually impaired [2]. This visually impaired population is at a disadvantage when it comes to communicating since its members are not able to interpret most nonverbal messages [3]. Humans rely on visual cues when interacting within a social context and for those who cannot see, this interferes with the quality of social interactions [3].

Previous studies have shown that non-verbal communication is integral to maintaining social interactions. Since the majority of non-verbal communication is comprised of facial gestures and expressions, which are visual cues, interper-

sonal communication is compromised [3]. This societal disadvantage can best be addressed through an application of science and technology; we propose the implementation of computer vision to provide realtime feedback of facial identification and expression for the visually impaired.

Through this study, we seek to answer the following question: how can computer vision technology effectively convey nonverbal, and otherwise inaccessible, information to blind individuals in face-to-face interactions with sighted individuals? We believe that the use of computer vision to analyze and relay nonverbal messages in an interaction will improve the quality and ease of face-to-face communication for the visually impaired. Our use of computer vision involves developing an assistive device capable of analyzing faces, providing realtime feedback, and conveying the facial cues of those with whom the visually impaired interact. Idea generation and accuracy and subject testing of the software package can then be conducted through focus group studies and user feedback. In this thesis, we examine whether realtime feedback of facial expressions can allow visually impaired subjects to communicate more effectively, and we share our work with hope that it may inspire future research in the field of computer-based handicap assistance.

1.0.1 Visual Impairment

According to the World Health Organization, there are about 284 million people in the world who are visually impaired; 39 million of this population are blind [2]. In Maryland alone, the site of this study, there are over 100,000 individuals who are legally blind or experience total blindness [4]. Typically the main causes of visual impairment include refractive errors (myopia, hyperopia, astigmatism),

cataracts, and glaucoma. Although anyone may be at risk for visual impairment, the risk increases with age; nearly 65% of visually impaired individuals are 50 years old or older [4].

It is crucial to understand the nature and definition of various classifications of blindness before designing an assistive device. Currently there are four classifications of visual acquity: normal vision, moderate visual impairment, severe visual impairment, and blindness. Within the category of blindness, there are two main forms: congenitally blind and non-congenitally blind. Congenitally blind individuals are those that have been blind since birth or an illness shortly following birth, whereas noncongenitally blind individuals are those that became blind later in life due to a variety of factors including macular degeneration, strokes, etc.

For non-congenitally blind individuals, the Social Security Act defines the confirmation of legal blindness as those that have a visual acuity of 20/200 or less in both the right and left eyes, even when wearing glasses or contacts. In addition, if the overall field of vision is less than 20 degrees, the individual is considered to be legally blind. Approximately ten percent of those who are legally blind have absolutely no vision; the remainder have varying degrees of light perception. Those that have absolutely no vision are described as having total blindness. Despite the varying types of visual impairments, there are currently no differences between how the congenitally blind, noncongenitally blind, and sighted perceive emotion cues and expressions [5]. Thus, the prototype can be designed to recognize facial expressions in the same way for all users [6]. However, the prototype must allow a user interface through which a user can select whether facial expression and recognition are applicable in their everyday lives.

An assistive device designed for the visually impaired population must focus on age and technology proficiency segments. The device must be useable by both newcomers and those familiar with assistive technology. This requires a sophisticated algorithm which is user-friendly in both hardware and software design. In addition, the prototype should be widely applicable to visual impairments while especially targeting the older population segment.

1.0.2 Computer Vision

Computer vision is the machine extraction of information from visual feedback. Computer vision systems may be configured as mono-vision (one camera) or stereo-vision, which involves multiple cameras placed apart to allow distance and depth perception. Computer vision can be used in industry and quality assurance, medical image processing, and military and surveillance operations [7]. Currently, computer vision applications in facial recognition are limited to security applications. Since computer vision technologies have not typically been applied to handicap assistance, our study will provide insight on the feasibility of computer vision as an assistive technology. [8].

Further information on computer vision systems is detailed in the following Literature Review chapter, along with overviews of facial recognition, and expression recognition algorithms. The methodology of developing the design of hardware and software for the completed prototype, the surveys and interviews conducted for idea generation, and the experiments conducted for quality and accuracy testing are also detailed in the following chapters. Finally, this thesis is appended with transcripts and Internal Review Board documents detailing the particulars of testing procedures.

Chapter 2

Literature Review

The challenges that arise in designing a portable computer vision system include the need of a stable light source, the need for stability for cameras, power and processing requirements, and possibility of damage to the equipment. Current innovations in the field of computer vision yield many different solutions. The following projects all use unique combinations of hardware and software to develop a solution in an effort to address these concerns and the needs of the visually impaired community.

2.1 Past Computer Vision Systems

Gemstone Team Vision integrated a GPS chip, an inertial navigation unit, and a low-resolution USB webcam with audio and haptic feedback to provide an easy-to-use navigation and object recognition package for visually impaired individuals. Team Vision mounted the USB camera onto the bridge of a pair of sunglasses, and wired the system to a laptop computer running Windows XP [9].

In 2008, McDaniel et al. [10] created a haptic belt to convey the location of a person relative to the user's field of view using vibrotactile signals. This type of feedback is more effective than auditory cues in communicating location because

it takes advantage of the sensitivity and reliability of a visually impaired individuals sense of touch. In addition, vibrotactile signals are not open to variable interpretation and do not interfere with other activities of the visually impaired.

Researchers from Arizona State University, created a facial recognition prototype using a pinhole camera in a pair of glasses for their system [3]. The analog CCD camera took video that was converted in the digitizer, which transferred the video into a digital format that could be interpreted by a laptop computer. When the computer was able to recognize a face and put a name to that face, Microsoft Speech Engine was able to give the user verbal feedback, converted from text on the laptop. In their conclusion, Krishna, et. al. [3] discussed the potential for future work using a camera of better quality and alternative computing elements such as PDAs or other handheld devices.

Each and every one of the aforementioned studies have used a single camera for image acquisition. However, stereo vision has been used in other research to provide better accuracy and versatility when gathering facial data as compared to monocular vision. This method allows for 3-D face tracking and recognition of faces and expressions with improved robustness than a single camera setup, and has not yet been applied in an assistive device for the visually impaired. Stereo vision can also help to stabilize a video feed that comes from a camera that is not mounted on a tripod [11].

2.2 Prototype Design

While creating different potential prototypes, we looked at the strengths and weaknesses of prototypes that were the focus of other computer vision research in recent years.

In 2005, Krishna et al. [3] developed a wearable face recognition system. They used a pinhole aperture analog CCD camera mounted in a pair of sunglasses. The analog feed was then converted to digital video format through an Adaptec video digitizer. A tablet PC with an Intel processor was used to process the images and implement a text-to-speech converter to inform the user of positively matched faces [3].

Treuillet and Royers system utilized a single body-mounted camera, as the main focus of their prototype [12]. Although these researchers focused on navigation using global positioning system (GPS) sensors, their prototype is still a very helpful concept. In creating their prototype, the researchers aimed to create a device that would be wearable, inexpensive, and mobile. Inspiration for their design came from wheeled robots, capable of autonomous navigation. The camera used in this prototype was an AVT Guppy 044B camera with a 2.8 mm lens. A body mount held the camera at a diagonal angle, allowing it a 120-degree view field, and connected it to a computer, located in a backpack. The body mount positioned the camera in the center of the body over the ribcage. In the conclusion and future work section, the researchers shared their plan to test a second prototype design, which includes a camera mounted on eyeglasses.

Johnson and Higgin's system transformed information received through the video camera back to the user through physical tactics [13]. Their system provided tactile feedback to the abdomen for the systems user. The video is taken through two webcams attached to a belt on a firm platform. There is a fixed angle between the cameras, with their lines of vision converging at ten feet in front of the user. The researchers specified that their materials entirely consisted of off the shelf components. The tactile feedback portion of the system is made

of several small vibrating motors. In their discussion, the researchers explained a plan to overhaul the tactile feedback system, replacing it with VLSI vision chips and an electro-tactile array, which will be less cumbersome and power-hungry. However, there is no mention of modifying the video platform for the prototype, despite their use of inexpensive webcams as a method of obtaining video.

While there are advantages and disadvantages of every system, we took all of them into consideration as we designed our own computer vision system. In addition, we obtained feedback from the visually impaired to make a more realistic system.

2.3 Facial Recognition

In the last twenty years, the computer-based facial recognition field has expanded rapidly. Several algorithms have been introduced and improved to the point where computers can rival humans in accuracy of facial recognition [14]. In order to develop our product, we need to understand how we identify faces, and to understand and evaluate the different existing facial recognition algorithms and examine existing applications of this technology.

Sinha, Balas, Ostrovsky, and Russell outline nineteen basic results regarding human facial recognition, including many of the methods that humans use to identify faces [6]. They show that the study of human processes involved in facial recognition and the artificial algorithms being used for facial recognition systems are inextricably linked together. The human brain can recognize faces in 120 milliseconds (ms). In order to achieve a useful system, the algorithm we choose must have near realtime feedback.

Several real time algorithms have been developed in recent years. Ross Bev-

eridge of the Colorado State University evaluated the efficiency and accuracy of the algorithms, Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Elastic Graph Matching (EGM), and Bayesian Intrapersonal/Extrapersonal Image Difference Classifier (BIC) [15]. A study by Krishna, Little, Black, and Panchanathan also evaluated these algorithms with respect to changes in illumination and pose. The LDA and PCA algorithms were found to be superior. LDA was fastest while PCA was the most accurate [3]. We will be evaluating the four algorithms (PCA, LDA, EGM, and BIC) to determine what works best for our purposes, as our conditions may be slightly different than those of previous studies. Another algorithm recently developed is Luxands FaceSDK, a facial recognition system that supports recognition for still images and real time video streams. It is mostly used in the entertainment industry to create real time animations.

As outlined by Zhao, Chellappa, Phillips, and Rosenfeld [16], these algorithms may be susceptible to several well known challenges including pose, illumination, and resolution. However, over the past several years, major improvements have been made to these baseline algorithms. In an experiment by Alice J. O'Toole, seven facial recognition algorithms were compared with humans on face matching tasks. Out of the seven algorithms, three were better at recognizing faces than humans were [14]. Though illumination still presented problems, O'Toole's study shows that current algorithm capabilities compete favorably with human ability to recognize faces.

Another challenge to consider is the movement of visually impaired subjects, which may cause the acquired faces to be blurred. We need to develop image deblurring tools and feed the restored face images to the recognition algorithms.

Systems for recognizing faces do currently exist. However, many of the existing facial recognition systems are created for security rather than for the visually impaired [17]. Even so, these systems show that it is possible for a recognition system to recognize faces acquired under controlled conditions at a recognition rate of 99.2% in near realtime. One system, developed by Krishna et al., uses a PCA algorithm and was designed with the visually impaired in mind [3]. Krishna's system does not use stereo cameras, nor was it tested on visually impaired users. Krishna's system also did not have a facial expression recognition feature. Our system will be designed with the advice of the visually impaired to provide a facial recognition system that would benefit them the most.

2.4 Expression Recognition

In order to use facial expression recognition algorithms to improve social interactions among the visually impaired, we need to understand the biological underpinnings and universality of facial expression, as well as research existing algorithms.

Paul Ekman and Wallace V. Friesen [18] identified six universal facial expressions in an experiment conducted among a group of New Guinea tribe members who had been isolated from foreign cultures. They found that the subjects were able to accurately identify certain expressions happiness, anger, sadness, disgust, surprise and fear depicted in pictures of different facial poses. Their findings support the theory of the universal correspondence between certain facial behaviors with certain emotions [18]. These expressions are characterized by the facial action coding system which uses 44 small facial movement action units (AU), each of which encodes a specific facial muscle action [18]. Since the expressions stated

above can be recognized cross-culturally, the primary focus of our computer algorithm will be on detecting these six universal expressions.

Experiments have also been conducted where motion in the major facial areas (the mouth, eyes, and eyebrows) was selectively frozen through videos. Multiple studies concluded that in some expressions, single movements could play the decisive role in differentiating facial expressions while, in other expressions, multiple movements were necessary in differentiating facial expressions [19].

Takeo Kanade, Jeffrey F. Cohn, and Yingli Tian compiled a database of 2105 coded facial expression images using FACS [20]. This database consists of thousands of different faces, all with various expressions, poses, and illuminations. We will test the chosen algorithms and final system on this facial database to ensure accuracy and efficiency.

One prominent expression detector is a parametric optical flow-based algorithm that analyzes facial expressions in realtime. It was developed by Yacoob and Black in 1997. This algorithm employs facial tracking and facial feature tracking, checking changes against thresholds established by existing databases [21].

Another expression algorithm that is available in the market is FaceAPI, a face tracking technology developed by Seeing Machines Limited. FaceAPI provides real time face recognition and facial feature tracking by tracking X individual points on the face in the xyz plane. The algorithm is able to detect head position and has a freedom of +/- 90 degrees for head orientation, allowing more freedom for the user of the camera to move around during the conversation. FaceAPI is also able to calculate distance of a person without the use of dual cameras (a stereo system), but rather by calculating the size of the face and determining the

distance using that measurement.

Having code to detect different expressions is essential to creating a feedback system, which shares such information with blind users. Shafiq ur Rehman at Umea University in Sweden wrote his doctoral thesis on emotion recognition feedback through vibration [22]. After focusing on the importance of recognizing and understanding emotions in everyday conversation, Rehman proposed a method to communicate such information without vision. He created a chair with what he calls a vibration actuators matrix located on the back of the chair. Distinct sequences of vibrations represent different emotions, and the user therefore knows which emotions their communication partner is exhibiting through memorization of the vibration pattern. Rehman believes in the value of tactile perception and incorporates his feedback systems mainly with tactile functions such as vibrations.

To make our system reliable, we need to pick and develop algorithms best suited for face and expression recognition. The method in which the system communicates with the user is also important. Unlike many of the other studies, we will implement a design with the preferences provided by blind individuals.

Chapter 3

Methodology

Our team of researchers was composed of undergraduates at the University of Maryland, College Park, enrolled in the Gemstone Honors Program. Gemstone is a four year research program that selects students based on academic excellence and interest in research and forms multi-disciplinary teams that direct and conduct research. Each team includes faculty mentors to assist research and development during the four undergraduate years, and a number of experts in the field to vet the research conducted. Due in large part to the support and funding provided by the Gemstone Honors Program, we were able to choose a Product Design cycle to develop a computer vision device for the visually impaired.

A product cycle consists of four phases. The Idea Generation phase requires internal review of ideas and past literature, and external review of designs, specifically through survey/interviews. Following idea generation, the Design and Prototyping phases involve the design of software, hardware, and the overall device. The Troubleshooting or Evaluation phase then allows for review of designs and prototypes before returning to further idea generation.

To outline our methodology, we first brainstormed possible basic ideas for our code and for product design based on the literature review and opinions from

potential blind users. Once we decided upon a basic design of our prototype, we continued our surveys and interviews. During this phase we communicated with both sighted and visually impaired people to observe the social response to our prototype and gather data to help us improve the social acceptability of our prototype design. We then proceeded to the product design phase, where we took into account the results from the surveys and interviews. After developing the software for face and expression recognition and finalizing a hardware design, we physically produced the prototype using the available resources and materials at hand. Finally, we moved on to evaluation, where we conducted subject testing with visually impaired users to test our device in a controlled environment. After obtaining data on the accuracy of the system and feedback from the test subjects, we analyzed the feedback for future revision. The cycle then restarted, and we began again with idea generation to redesign our system.

After examining the literature, including the work of a former Gemstone team, Team Vision, we formulated several ideas for mounting the hardware so that it worked with our algorithms and is convenient for the user. Following an internal review and agreement on several different components for the final device, these ideas were put to members of the Technological Evaluation Team at the National Federation for the Blind (NFB). As part of the National Federation of the Blind, this team is well informed on current technological advancements in devices for the visually impaired. By evaluating our designs with the Technological Team, the device is uniquely developed for the user.

An initial idea for the camera mount was to attach a small camera on a pair of sunglasses, as done for other assistive devices for the Blind [9] [3]. This would put the camera at head height, which is ideal for both recognition and expression

analysis algorithms. However, the problem with this approach was the size of the cameras needed for our application. In order to have the needed resolution, the camera would be very visible on the glasses. We decided to purchase a remote head camera from Point Grey, so that the lens could be mounted on the glasses and the larger circuit board could be mounted elsewhere. These ideas and others were presented during our meeting with the NFB.

Several members of the Technological Team with limited visual impairment were concerned that a sunglasses mount would hinder their ability to use their limited sight. They were also concerned about the conspicuousness of wearing sunglasses all the time. A pendant concept was also reviewed and discarded as being too obtrusive and noticeable. As shown by researchers at Arizona State [23], negative attention due to "being different" is a main factor in whether an assistive device is used by the target population. Keeping this in mind, the mount selected as most unobtrusive and light was to have the camera mounted on a white cane, as suggested by members of the NFB.

The white cane idea seemed to be the most practical idea. The blind users already carry around a cane on a regular basis. This way, there is no need to carry an extra device. The top of the cane is generally at eye level, which makes orientation for the algorithms simple. Mounting on the cane also provides stability for the camera. The cables could run along the length of the cane and connect to the laptop relatively easily. For these reasons, we decided to pursue the white cane mounted camera design.

Chapter 4

Survey/Interviews

4.1 Blind interviews

Our main objective in interviewing blind individuals was to gain insight into their interactions with both blind and sighted members of society, and to garner feedback about specific aspects of our proposed device. While we realized the extent to which their societal experiences and opinions would differ, we wanted a more complete idea of the limitations that they have encountered in societal interactions to better tailor our device towards their needs. We interviewed nine participants from the Columbia Lighthouse for the Blind, the University of Maryland, National Federation of the Blind, and other associates. Of the nine individuals, eight were completely blind and one had light perception. All interviewees were between the ages of 40 to 70 years. The first question asked the participants to describe their social interactions with both blind and sighted people. All of them answered that they had experienced no problems interacting with blind people, with one individual saying "because we can identify with someone with [the] same disability." However, their experiences and comfort interacting with sighted individuals varied greatly. Some claimed that it was no different than talking to

a blind individual, whereas others cited that the lack of certain factors, such as eye contact and facial expression, detracted from their conversations.

Interestingly, although these variables are also absent during blind-blind interaction, participants only cited these as causing problems in blind-sighted interaction. (In the case of eye contact, this may be due to the blind individual's heightened self-consciousness and/or the sighted individual's unease). One participant stated that the sighted are "uncomfortable with the lack of eye contact. . . so that's an issue." Another participant explained that, while the blind have learned to recognize the nuances of tone to deduce the emotional state of sighted conversational partners, "it [would be] good to know before you talk to them, what they have on their face."

In order to assess the current assistive technologies available to the blind, we then asked participants to describe any assistive devices that they used in social interactions. Several indicated that they used the white cane for mobility purposes. A couple said that they used the Job Access With Speech (JAWS) software program, designed to help them navigate the internet and computer applications independently by reading aloud everything on the screen. Other devices that the participants use include a note taker (e.g. daily planner), digital recorder, Braille display, and gadgets (e.g. cell phone, watch, calculator) that speak out loud.

The third question asked participants to rate how important it was for them to be able to recognize emotional cues in their daily social interactions. On a scale of one through five, one being the lowest score and five being the highest, five participants responded with an answer of four or above, three responded with an answer of either 3 or 3.5, and only one responded with an answer of

between one or two. The individuals who gave a rating of four or five said that being able to recognize emotional cues would enhance intercommunication and personal safety. Those who gave a rating of three or below claimed that being able to recognize emotional cues was not very important to them since they have already lived without the ability to do so for so long.

In our series of questions, we asked participants to rate their interest in a portable device that could convey facial identities versus one that could convey facial expressions. The majority of participants was enthusiastic about the idea of being able to perform either one of these tasks, but when asked to give their preference, most showed a stronger interest in expression analysis over identity analysis. One of the participants stated that being able to "pick up on these cues ... could be a matter of safety." Another participant pointed out the advantage of being able to recognize facial expressions on any individual versus only being able to recognize the facial identity of an individual whom the blind user already knew (and had input into the database). As another interviewee explained, while blind individuals are able to discern certain emotional cues based on the voice inflections of someone they are familiar with, they cannot recognize the emotional cues of someone whom they do not regularly converse with because they are not familiar with the individual's voice signature; thus, a device with the ability to convey the facial expression of this individual would be useful in such situations. The only participant who expressed a stronger interest in identity analysis said, "since I've never had the ability to [recognize emotion] ... visually, it really doesn't matter all that much. I'd rather have facial recognition."

One particular interviewee who lost his sight later in life explained that the nature of his blindness contributed to his stronger interest in expression analysis.

He claimed that "people like me who've seen... for facial recognition, that's a curiosity of totally blind people [people who were born blind], to know what your faces look like. For me, it's more the expression of their feeling because I know what their face looks like." While not enough information was collected on the nature of our participants' blind conditions to assess the validity of this statement in our first round of interviews, we want to explore his contention further in future interviews.

Thus, from our interviews, we have gained more insight into the current state of interactions between the blind and sighted, and the idealized state of societal interactions from blind individuals' perspectives. Some blind say that their interactions with the sighted are no different from their interactions with other blind; on the other hand, others explain that interactions with the sighted are sometimes uncomfortable and/or lacking certain factors. While we cannot restore their vision, we seek to enhance the blinds societal interactions by enabling them to recognize facial identities and expressions, like the sighted, with our proposed device. Ultimately, we want to combine the identity and expression analyses functions into one portable device, so that blind users have the option of using one or both to their preference.

4.1.1 Results

The participants of the interviews hailed from a variety of backgrounds and had varying degrees of visual impairment. The nature of the visual impairment becomes apparent in the preferences of congenitally blind, and noncongenitally blind individuals.

- Interviewee 1: congenitally blind, ranked identities higher

- Interviewee 2: nature of blindness unknown, ranked expressions and identity equally
- Interviewee 3: lost sight later in life, equally enthusiastic about both
- Interviewee 5: lost sight later in life, ranked expressions higher
- Interviewee 6: lost sight later in life, more interested in expressions
- Interviewee 7: congenitally blind, ranked expressions higher
- Interviewee 8: light perception, ranked expression higher
- Interviewee 9: lost sight later in life, would prefer expression

Congenitally blind individuals enthusiastically hail both the identity and expression recognition systems, whereas participants who lost their sight later in life have a greater preference for the expression recognition system.

In response to the importance of facial recognition technology, users shared the following insights:

- "Being able to recognize faces and [interact with] them and know who they were based on just that real quick "Hi Jane," without them telling me who they are. That would be more important to me."
- "That would be really cool. Then you could interact with them. You would know who they were and maybe theyd be somebody youd want to talk to."

In response to the importance of facial expression recognition technology, users shared the following insights:

- "Like I said, it is very important. Especially if we don't have the visual acuity to really see the facial expressions . . . so in order for us to be able to comprehend things, make good decisions and live life to the fullest, it's important if we can't see it, we should be able to feel it or hear it or experience it through a different outlet just like someone's sight."

- "You'd go ahead and say, "Hello, how you doing?" or you ask them a question, and say they're in a very depressed state . . . we've got to be able to pick up on these cues. It could be a matter of safety. Safety should be number one in everyone's case."

- "You don't know if someone is upset by what you're saying or reacting negatively. You could just make the situation worse by not recognizing that and changing your communications accordingly."

- "I think it's important to understand where people are coming from, so you can have a better connection with peopleif you can sense, if you know they're sad, or happy, or uncertain, or troubled, or whatever."

- "I lost my vision 14 years ago. I miss seeing people's faces. To see if they are angry, happy, or disgusted. I don't care about the color of the person or if they are attractive or not. It used to make a big difference to me. I want to see if what I said upset them or made them happy or sad. Sometimes I can tell with tones but not all the time. But anybody can fool anybody."

4.2 Sighted Surveys

We conducted online surveys with sighted participants to gain an understanding of how the sighted community would perceive blind individuals using our device. The sighted communitys opinions of our assistive device are relevant for two major reasons: (1) they influence the blinds comfort with utilizing our device, and (2) they affect the behaviors of both the blind and sighted engaged in interaction. By gathering feedback from sighted individuals, we aim to develop a prototype that fits the use for blind-to-sighted communication in addition to blind-to-blind interactions.

Survey participants were recruited via email listservs at the University of Maryland as well as emails to personal contacts. Responses were collected from a total of 98 participants ranging in age from 18 to 63. The mean and median ages lie within the 20-29 range. 36 respondents (36.7%) were male while 62 (63.3%) were female. Our sample does not accurately represent the general population, but we believe that the responses can nevertheless prove insightful for recognizing basic trends in attitudes towards the proposed device.

The first question of our survey asked participants to identify the frequency of any interactions they have or have had with the blind. Possible answer selections were "never," "rarely" (about once every five years), "sometimes" (about once a year), and "often" (at least once a month). The frequency distribution was heavily skewed towards fewer interactions with the blind. Thirty four percent of participants claimed to "never" interact with the blind while forty seven percent did so "rarely." "Sometimes" and "often" were selected by fifteen and four percent of participants, respectively.

When asked if they thought there were "any societal stigmas associated with

the blind," 73 sighted participants (74.5%) said yes, 22 (22.4%) said no, and 3 (3.1%) said the question was not applicable to them.

Of the participants who believed there were societal stigmas, 28 provided comments centered on the theme that blind people are seen as incapable of taking care of themselves fully. For some, this meant that the blind could even come across as a "burden" or "unemployable." 18 participants expressed the belief that most people tend to ignore the blind because they are uncertain of how to behave in their presence. The fear of being offensive and/or feeling awkwardness is seen as a reason for distancing oneself from the blind. Five sighted participants went as far as to suggest that society associates blindness with partial mental and social deficiencies. The sighted community's perception of the visually impaired and blind is not well-documented in the literature and, therefore, we hope that our survey findings can offer some additional insight on this topic.

Next, survey participants answered a two-part question regarding their level of comfort in response to hypothetical events involving our proposed device. The first part asked, "How would you react if you saw a person carrying a cane mounted with a camera, which connects to a laptop in a backpack?" Seventy-nine percent of participants responded that they would be "intrigued" by someone carrying such a device. Eleven percent selected that they would be "initially uncomfortable," indicating that their opinion might change with time and familiarity. "Uncomfortable" and "neutral" respondents comprised four percent and six percent, respectively.

Then, the second part of the question asked, "How would you react if you knew that such a device was analyzing your facial expressions and identifying your face?" For this, fifty-two percent of participants selected "intrigued" while

thirty percent picked "initially uncomfortable." Ten percent identified with "uncomfortable," and the remaining eight percent selected "neutral." The shift in the data between these two questions depict a society that is interested in and open to general assistive technologies for the blind, but the introduction of facial expression and identity recognition capabilities correlates with increased caution.

The final survey question asked, "Would such a device negatively affect any interactions you may have with the blind?" Due to the vagueness of the word "negative," we accepted open-ended responses instead of posing multiple choice selections. The majority, 65 percent, of participants made clear that they did not think the device would adversely affect any interactions they might have with the blind in the future. Five percent predicted that it would, however. The remaining thirty percent of participants gave responses that we categorized as "others." These participants were either unclear or unsure of how their behavior would be influenced. Many of those who said the device would bear no negative influence as well as those who fell into the "others" category expressed the desire to learn more about the device, suggesting that their feelings would likely improve with increased knowledge and familiarity. On the other side of the spectrum were those who expressed unease and suspicion. Several asked whether the device would store information about their faces without their knowledge or consent. Others questioned the accuracy and reliability of the device.

For every question in the survey, we allowed the sighted participants to expand on their responses. The following are prevalent themes among the open-ended responses. Most of the sighted individuals who had expressed comfort with and even support for our project were hopeful of its potential benefits for the blind. One participant wrote, "I would be fascinated and would support any device that

might aid the blind." "[The application of the device] would be similar to another person looking at me and analyzing my facial expressions with his or her sight and trying to figure out my emotions or opinions," wrote another sighted individual. On the other hand, several participants argued that a lack of knowledge about the device could lead to initial and/or long term unease and concerns about privacy invasion and potential computer misinterpretation. One wrote, "[It is] always a bit disconcerting to feel like you're being recorded without permission." "People don't even interpret others' expressions properly so it would be hard for a machine to," said another. Contrary to their suspicions, our device will not save any facial (or any other private) information. A third group of participants pointed out that the presence of the device might affect the way they behave. One participant replied, "It might make me more self conscious about my mannerisms than I otherwise would be." Another said, "I may be more aware of how I act towards [the blind user], and my expressions may not be genuine, for better or for worse."

4.2.1 Results

While we understand that our ad hoc survey respondents are not representative of greater society, we still find it important to note how seldom sighted individuals interact with the blind individuals. As shown, the frequency of interactions is low; it is unclear however whether this is due to the lower number of visually impaired individuals encountered by our respondents or due to communication difficulties precluding visually impaired individuals from creating a greater societal presence.

Overall, the survey results show that many sighted individuals would be comfortable with the blind using our assistive device, if not immediately, then with

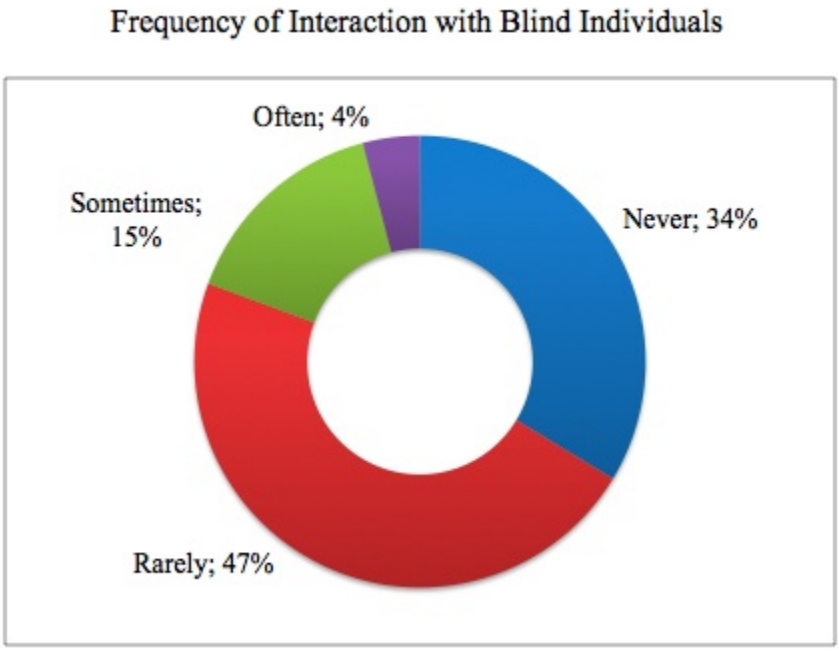


Figure 4.1: In this graph, "rarely" represents interacting with blind individuals approximately once every five years, "sometimes" represents approximately once every year, and "often" represents at least once a month. The data are in accordance with our hypothesis that visual impairment creates a barrier between mainstream society and the blind community.

time and familiarity. However, when not clearly understood, a system that tracks and analyzes facial features will raise concerns of privacy invasion and misinterpretations caused by computer inaccuracy. Our third finding is that a sighted persons awareness and knowledge of the device can cause a reflexive alteration in their behavior. The social stigma associated with assistive technology can only be decreased with increased awareness of technology and the equality which it can provide.

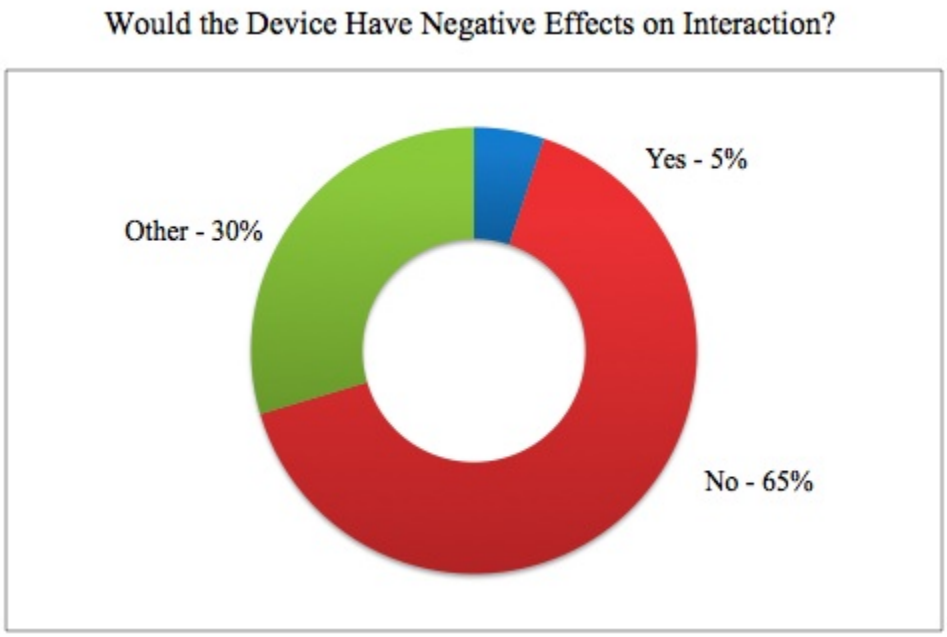


Figure 4.2: The majority of sighted participants in our survey believed that our device would not have negative implications on interactions with blind individuals. Those who responded with "other" thought that there would be no negative effects on interaction as long as they knew what the device was doing, and provided that the device was not storing images of their faces.

4.3 Survey on User Preferences

Our main objective in conducting an online survey among blind individuals is to gain further insight into whether there is a correlation between onset of blindness and preference for a specific aspect of our prototype (facial or expression recognition).

Through our pre-prototype blind interviews and post-prototype blind experiments, we have found that all of the noncongenitally blind participants preferred the facial expression recognition aspect of our prototype, while those who were

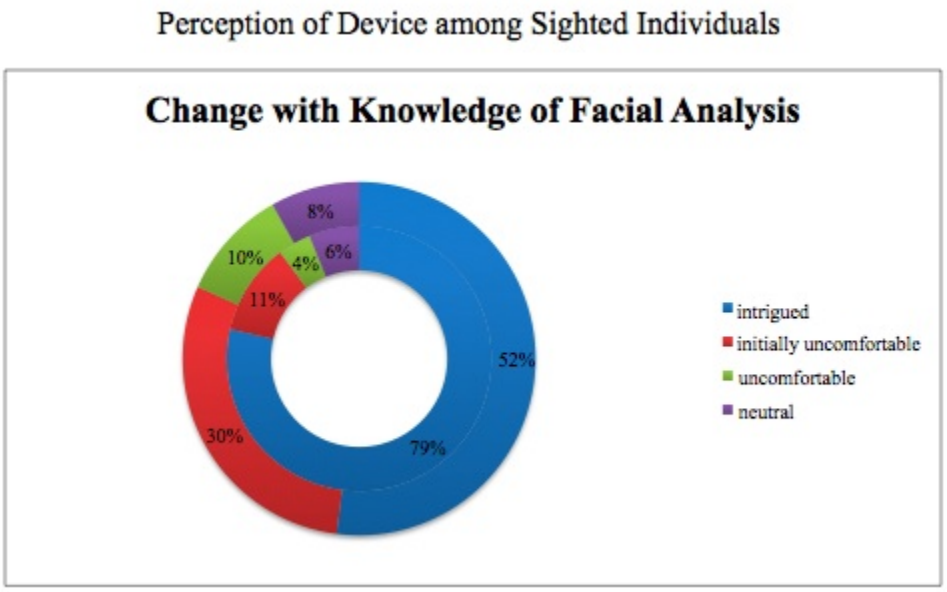


Figure 4.3: Society seems to be open to the general idea of assistive devices for blind people. However, the novelty of our device is accompanied by unfamiliarity and a shift in attitude. Those who indicated that they would be "initially uncomfortable" attested that they would probably become more comfortable given more information and time.

congenitally blind were enthusiastic about both the facial recognition and facial expression aspects of our prototype. Here, facial recognition is defined as the ability to identify a face and match the face with a name, whereas expression recognition is defined as the ability to see faces that convey happiness, sadness, surprise, etc. Though our sample size was relatively small, with nine interviewees and five experimental participants, the preferences have been in unison. We hypothesize that the congenitally blind are enthusiastic about both facial recognition and expression, as they have never been exposed to either ability and thus would like to experience both characteristics of the sighted. Furthermore, we

hypothesize that the noncongenitally blind prefer expression recognition because they have been exposed to expression at birth and thus have become more reliant on the ability to see expressions to gauge emotions.

We generated a hypothesis based on the results of our previous interviews and experiments to test on a greater sample and wider audience of blind participants through an online survey. After obtaining IRB approval, we sent out the online survey to recruit as many blind participants as possible from the Columbia Lighthouse for the Blind, the University of Maryland, National Federation of the Blind, and other associates. Questions were targeted at determining the correlation between time of blindness and technical preference. Thus, questions asked the blind participant to indicate the time at which they lost their sight, the extent of their blindness (i.e. ability to sense color, light, see outlines of objects, etc), and preferred recognition (expression or facial). We also included some other questions regarding their comfort level of using the device, so that we could identify potential confounding variables in our data set. Therefore, questions also asked the blind participant to indicate how often they use assistive technology and what they primarily use the technology for (i.e. navigation, object recognition, etc).

We had 40 participants, 20 who were congenitally blind and 20 who developed blindness later in life, were noncongenitally blind. As predicted, congenitally blind participants responded that they preferred the expression recognition and facial recognition technologies equally. Participants who developed blindness later in life generally preferred facial recognition over expression recognition technology. Due to a small sample size, however, we cannot claim that this presents a significant relationship. Further studies are required to formulate a definitive conclusion as to whether there is a correlation between the nature of blindness

and preference for expression recognition.

The following table relates the percent of our participants that reported familiarity with a variety of assistive devices, promoting the need to further studies in the usage of assistive technologies and how they impact social interactions.

Table 4.1: Assistive Technologies used by Respondents

Assistive Technology	Congenital Blindness	Acquired Blindness	Total Users	Percent of Respondents
Communication	18	21	39	95.10%
Object Recognition	7	11	18	43.90%
GPS Navigation	4	3	7	17.10%

In team experiences of conducting subject tests and conversing with blind individuals, we observed the tendency of the sighted team members to revert to non-verbal communication cues such as nodding, followed by a quick correction to ensure the blind participant received a response. Another hypothesis that can be explored in further studies is that with a device translating non-verbal cues, sighted individuals might find greater comfort in conversing as accustomed.

Chapter 5

Product Design

The ease-of-use, effectiveness, and reliability of the white cane has led to its ubiquity in the visually impaired community. Our device is designed to supplement the white cane carried by many of our target population. Mounting the camera on a white cane decreases bulk and the need for users to continuously maneuver and adjust. It is also a stable platform around eye level for most users and eliminates the need for the user to carry another device, as they are already carrying a white cane. The mount itself must be sturdy and must not upset the balance of the white cane as a user walks with. In addition, it should be easily removable when not in use, and re-attachable without difficulty.

5.1 Previous designs

The original prototype idea involved mounting the camera to a pair glasses, similar to Krishna's design [23]. While sunglasses are commonly worn by blind individuals, those with limited vision prefer not to obstruct their sight. Having met with such individuals, we discarded the idea of mounting a camera to glasses. In addition, the weight of a camera with adequate power for our application would cause it to be heavy and uncomfortable on glasses. Glasses also increase the

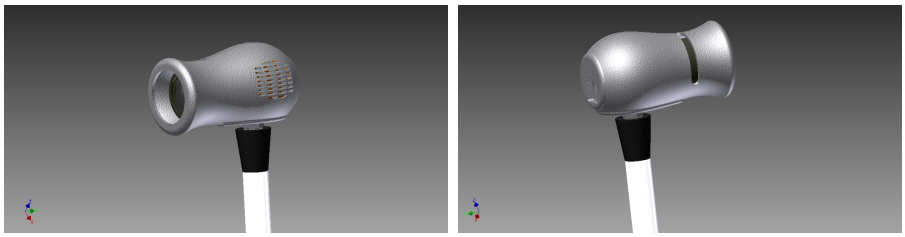
obtrusiveness of the device, and will possibly draw greater attention to the use of assistive technology.

To increase mounting options for users, a pendant design was created where the camera would hang around the user's neck. However, this does not control for the height of people a user might be interacting with, and requires the user to adjust for height. It would also present difficulties with aiming the camera and the stability of the image.

A potential idea for mounting on the torso was to mounting the camera to a shoulder or chest strap. The chest is stable and allows for a more complicated camera setup. However, the chest is not at eye level and people of different heights may experience difficulty with capturing a face. Although a chest mount decreases the range of movement allowed to the camera, the mount can be designed to be attached to various mounts, including a cane and the shoulders or chest.

Another issue in creating the mount is how the camera should be mounted and supported. We considered attaching a wooden disk to the top of the cane to house the camera and associated hardware, foam spheres to provide security and support, and a standard 4-20 camera mounting screw so it can be taken off the cane and screwed onto a standard tripod for use while sitting.

The type and number of cameras used in the system is a major factor in the design. We considered using two cameras mounted a specified distance apart to implement stereo vision. This presented a problem with the size of the device become too large and heavy to be unobtrusive. Using a remote head camera (the Point Grey Dragonfly 2) could alleviate some of these issues. The small size allowed us greater flexibility in the location of the camera, but created difficulties due to the size of the associated motherboard and the optical resolution of the



(a) Front View of Mount (b) Back View

Figure 5.1: The casing for the final mount prototype consists of a plastic shell which clicks together, providing support, safety, and venting to the camera

camera. Due to the smaller size, the resolution of the camera was on the lower end. The remote head proved to be difficult because of the limitations of ribbon cable length and the inflexibility of the ribbon cable.

5.2 Final Design

For the final chosen design, we mounted a camera to the top of a white cane. The cane is also marked with a raised bump on the side of the cane which corresponds to the front of the camera. The blind are still able to use the cane for its original purpose and can also use our device, eliminating the need to carry another piece of equipment. The rigid white cane we used also places the camera at approximately eye level and provides a stable platform for the camera.

The camera selected for the final design was a Flea2, also from Point Grey. This camera is fully integrated. It is more stable and has a better optical resolution than the Dragonfly2. It has a Firewire connection to the computer. This camera also comes with a mounting platform for a tripod screw /citepointgrey12.

The camera is mounted to the cane with a standard 4-20 tripod screw. This

allows the camera to be easily attached and removed from the cane. The camera can also be removed from the cane and placed on any standard tripod. This allows the user to place the camera on a table for use in a different setting.

In order to protect and disguise the camera, there is a plastic shell surrounding the camera. The plastic is lightweight and aesthetically pleasing with a vent on the side to prevent the camera from overheating. The shell is a snap fit design so that the shell may be removed to access the camera. The shell was created using rapid prototyping with ABS plastic. The final shell for the device can be fabricated through injection molding with an impact resistant plastic to better protect the camera.

The camera is connected with a Firewire cable to the laptop. A nine-pin Firewire cable was chosen due to its speed and flexibility. A locking cable from PointGrey was used to prevent the cable from being pulled out due to motion of the camera. The laptop is housed in a mesh backpack with some measure of insulation and padding. This protects the laptop from possible damage, but also allows the heat to escape, since the computer algorithms are processor intensive.

5.3 Prototype testing

The prototype was tested in experiments with blind participants, selected to test the device in its entirety. The participants were invited to familiarize themselves with the device and provide opinions on the balance, ease of use, and simplicity of the design. The blind participants were able to accurately point the cane-mounted camera at a person using the raised dot. The prototype was also tested with NFB participants. They found very little difficulty using the white cane for walking with the camera attached, with a negligible change in the balance

Figure 5.2: The final prototype consists of a digital camera screw mount on a white cane and a plastic shell casing. This mount is easily detachable and can be attached to a camera tripod as well

of heavier canes. There were some concerns about the distribution of weight, though all participants reported that it would not dissuade them from using the device.

5.4 Future work

Re-prototyping the plastic casing requires investigation of the casing, with possibly adding more layers of protection. Plastic injection molds must also be designed and the process evaluated. Different plastic materials would also be carefully considered for their ability to survive a fall. The mount must also be able to attach to a variety of canes, whereas it currently is only used with one rigid cane. Canes are selected based on width, and can be foldable, telescoping, and adjusted for height, as a matter of user preference. The camera must be able to integrate with all of these types of canes. The mechanism with which the

camera attaches to the cane could also be redesigned. By adding a clip which allows the camera to attach to a backpack/shoulder strap in addition to the cane increases ease of movement and gives users freedom in selecting how to use the device. Adding user controls to the cane, in addition to voice-activated selection of user preferences, can allow the user to turn the device on/off when desired. A button could be implemented in place of voice commands in order to make the device less obtrusive. There is also a possibility of including a light on the camera to provide illumination of the subject, which will increase the accuracy of expression and recognition. This could make the device more obvious and merits further consideration.

Chapter 6

Software Design

6.1 Face Recognition

One major problem our device addresses is that of face recognition. When a person walks into a room, a blind user should be able to use the device to find the identity of that person. In order to do this, we had to attempt and evaluate a few different facial recognition methods.

6.1.1 Choosing an Algorithm

Method 1: OpenCV and Principal Components Analysis

The first method we attempted to use was that of Principal Components Analysis, or PCA [24]. PCA is used in many areas of computer vision, but the idea behind it is to reduce a large dimensional space, for example, the space of many pixels in an image of a face, into a smaller dimensional space where comparison is more economical. In the context of face recognition, PCA takes an image of a given query face, and represents it as a combination of "component faces," or eigenfaces generated from a training data set. These eigenfaces represent the principal components of the space of all faces. PCA has a database of faces for

which this representation has already been calculated. By comparing the query faces representation with the database faces' representations, PCA can return the face that best matches the query face.

PCA was fairly easy to implement, but there were also a few major drawbacks which ultimately led us to decide to look for other methods for face recognition. First, PCA has no way of telling if the image given to it is a face or not. To address this issue, we tried to use OpenCV to detect faces for us, since OpenCV has an implementation of a face detector. But in practice, we found this face detector to be slow and not as robust as we had liked. PCA was also heavily influenced by the illumination of the query face given. Depending on the way shadows fell on a person's face, PCA could identify them as different people, which was not acceptable for our purposes. Finally, for PCA to work, a face had to be given at a completely frontal view. For our system, we wanted faces to be recognized at many angles, so again, PCA was not acceptable. While we did not quantitatively test the PCA algorithm, it did not meet our qualitative standard, so we chose to look for other methods of face recognition.

Method 2: Local Binary Patterns

The next method we considered was that of Local Binary Patterns (LBP) [25]. The LBP method of face recognition divides the face into small regions, and then calculates a local binary pattern histogram for that region. These local binary patterns should be the same regardless of illumination conditions. The algorithm then takes these many LBP histograms, and combines them into a single multi-dimensional vector. To implement this method, we found an existing implementation of LBP code online. This code would take an input image and

output a vector, essentially computing the LBP for us. The problem then became how we could match or compare these vectors. Again, we made a database of known faces, and calculated the LBP vectors for each of these faces. A linear search using Euclidean distance as a metric would have worked as a method for comparing faces, but linear searching requires each query vector to be compared against every other vector. With a large database, linear searching is very slow. Also, because there could be noise in images, and the top result of a linear search could be incorrect, we wanted to get the top five matches as opposed to the top one. In other words, we wanted to find the five nearest neighbors to a certain query image. By finding more neighbors, we could return the identification which was the most common in the top five, making us more robust to noise, or misidentifications.

With these constraints in mind, we decided to use a Kd-tree [26] to facilitate a k-nearest neighbors search, with k equal to five. A Kd-tree is a data structure used to organize multi-dimensional data, perfect for the multi-dimensional vectors returned by the LBP algorithm. The Kd-tree structure makes returning nearest neighbors easy, since the data is already organized. If we gave the Kd-tree a single query point, it could quickly return the nearest neighbors, greatly improving speed over a linear search.

Unfortunately, in practice, we found LBP to be lacking. While it was fast, it was not nearly as accurate as we had hoped, and we had to search for another method.

Method 3: FaceSDK

After considering PCA and LBP, we found a commercial software development kit with its own proprietary face recognition method called FaceSDK. The performance seemed to be exactly what we needed. It was robust to illumination changes, seemed to work in real time, and correctly identified people at various poses. Instead of trying to develop our own algorithm, we decided to purchase FaceSDK and use their proprietary method. Using FaceSDK was not as simple as we thought it would be. While most of the methods for face recognition were documented, because we were using FaceAPI for feature tracking for facial expression recognition, we had to create additional methods to convert images from a FaceAPI format into one usable by FaceSDK. Once we accomplished this, we ran into further roadblocks. FaceSDK saves images as face templates which are then compared against. This meant that every time the device would start, every database image would have to be loaded into memory before it could be used. Finding face templates is a computationally intensive task, and the device could take minutes to get ready before being ready for use. To solve this problem, we decided to preprocess the images. Initially, we loaded each image into the program, and computed the face template each time the program was run. Instead of doing this, we created a pre-processing program, which took each image, calculated its face template, and saved that template to a file. When the device ran, it could then just load the face templates directly instead of calculating them every time.

Another difficulty we faced was the searching method used in FaceSDK. FaceSDK only allows for linear searching. This means for each query face, we had to calculate its face template and then compare it against each and every

face template in our database. While the demonstration application for FaceSDK worked quickly using this method, in practice in our device, we experienced 2-3 second pauses every time the device tried to recognize a face - the device was not running in real-time. Since real-time face recognition was one of the goals of our device, we had to address this problem. In order to do so, we decided to separate recognition into a separate process thread, taking advantage of the multi-core processor on our hardware. Through separating recognition to another thread, not only could we recognize faces without causing a freeze, but we could also run recognition more often.

Similar to LBP, even the proprietary method of FaceSDK was susceptible to misidentifications. Despite being forced to use a linear search which was slow, because of our decision to use threading, we also decided we could keep track of the top five results in recognition as opposed to only the top result. By using this k-nearest neighbors approach, we greatly reduced the number of misidentifications.

The end result was real-time, robust facial recognition. When our device recognizes a face in an image via FaceAPI, it performs facial recognition on that image every 30 frames. By doing so, we guarantee the user will receive an accurate result from facial recognition.

6.2 Software Topography

In its final state, the device software can be represented as in 6.1

We have two distinct parts of the program the application part, written in .NET c++, and the DLL, which is written in C++. The application handles displaying of data in the GUI and audio feedback, and the DLL contains the face

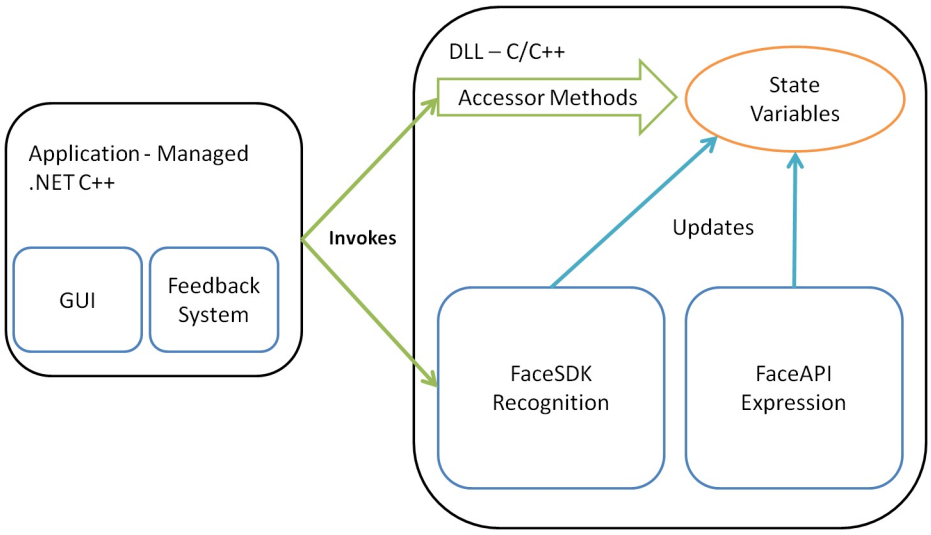


Figure 6.1: Software Overview

recognition and expression recognition code. The application can trigger these face recognition and expression recognition sections of code via public functions in the DLL. This relationship allows us to use voice commands to perform expression recognition and facial recognition. The DLL also has to pass back information to the application, which it does through getter functions. Whenever the application wants to know a persons expression, it must call one of these getter functions, and the DLL will return the identified expression to it.

6.2.1 Tracking Points on Faces

Using FaceAPIs ReceiveFaceData callback function in C++, we were able to read the coordinate information collected by the FaceAPI engine tracking the face within a series of images. We supplied the FaceAPI engine with a constant 30 frames per second stream of images collected directly from the Flea2 camera. FaceAPI Face-Texture coordinates were tracked for all 40 points shown in Fig. 6.2. As the program was executed, the coordinates were stored in memory and

were organized by Identification (ID) number. The points are described in Table 6.1

Table 6.1: FaceAPI face landmark standard

Landmark Identification Number Range	Group Description
0-99	Reference
100-199	Outer lip contour
200-299	Inner lip contour
300-399	Right eyebrow
400-499	Left eyebrow
500-599	Nose contour
600-699	Right eye
700-799	Left eye
800-899	Mask
900-999	Glasses frame (if wearing glasses)

For the purposes of determining the proper baselines and expression classifiers, 10 human subjects consisting of members from the research team (6 female, 4 male) were asked to carry out a simple expression procedure in front of the camera. Each subject was given the following instructions:

- Hold a neutral expression for up to 2 seconds
- Smile and hold for 1-5 seconds
- Return to a neutral expression and hold for up to 2 seconds
- Smile and hold for 1-5 seconds
- Return to a neutral expression

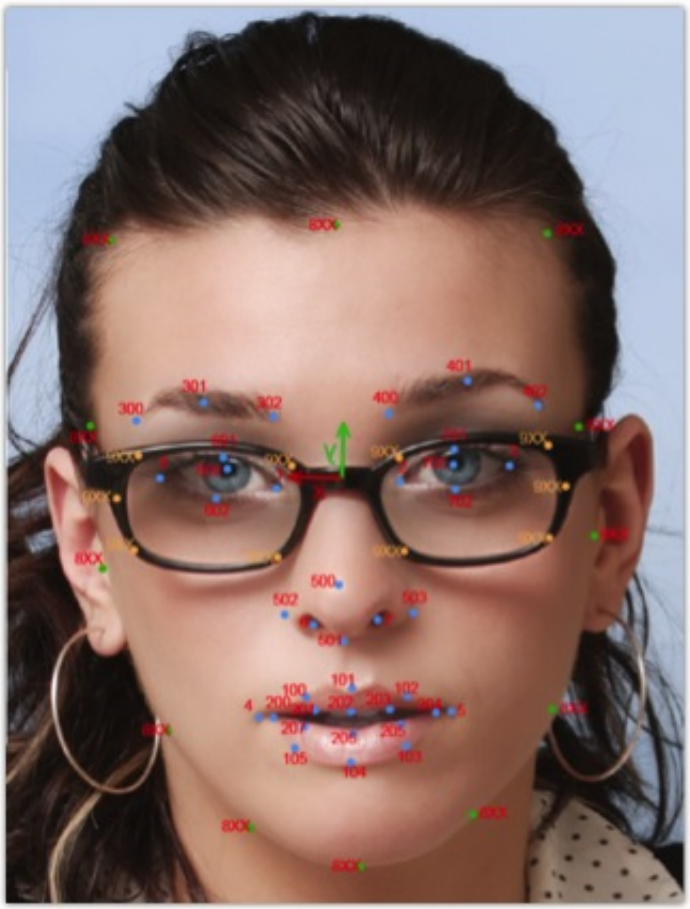


Figure 6.2: Points tracked on face. The numbers shown are the ID numbers that are referred to throughout this work

The instructions, specifically the definition of a smile, were left open to interpretation by each subject in order to account for any cultural or personal differences among the sample population. The Face-Texture coordinates were written to the computers Solid State Disk drive in a comma-separated value file.

After data acquisition was complete, the comma-separated value file was imported into Microsoft Excel for analysis. The x and y Face-Texture coordinate data for each of the 40 tracked points were plotted against frame numbers. Changes in the coordinate values were noted where there were changes in expres-

sion from neutral to smiling.

A similar approach was taken for the analysis of surprise, anger, fear, sadness, and disgust expressions. The same 10 human subjects were instructed to show the specified expression in the previously mentioned sequence. The x and y Face-Texture coordinate data for each of the 40 tracked points were plotted against frame numbers. Changes in the coordinate values were noted where there were changes in expression from a neutral state.

6.2.2 Expression Classification

Using the results of the tracking, we turned to classifying facial expressions in real-time. One classic expression detector is the parametric optical flow-based algorithm developed by Yacoob and Black in 1997 [21]. This computationally intensive algorithm tracks facial features in a sequence of images to detect the presence of expressions. Many traditional expression detectors, including Yacoob and Black's, have been too slow to be used for real-time detection. In an effort to increase usefulness, our device has been developed with a focus on speed.

From the baseline tracking, we determined that a consistent baseline could be calculated for each of the 40 tracked points by averaging the coordinates of each point as it appeared in the first two frames of a sequence. This is demonstrated in equation (6.1) below.

$$B[i] = \frac{fc[i][0] + fc[i][1]}{2} \tag{6.1}$$

Equation (6.1) calculates the baseline coordinate, B[i], for each landmark with IDnumber given by the integer i. fc[i][0] refers to the coordinate for landmark with ID number i at the first frame, and fc[i][1] refers to the coordinate for landmark with ID number i at the second frame of the sequence.

In the tracking data we observed several common trends among the expressions of different individuals. The data is discussed in more detail below. Here we used the trends to setup amplifiers for each of the six expressions studied.

Smile

As shown in Table 6.2, in smiles, 7 points contributed positively beyond the baseline, and 5 points contributed negatively beyond the baseline.

Table 6.2: Coordinate addition/subtraction arrangement for smiles

Description	ID#	Axis	Description	ID#	Axis
Left Corner	5	X	Bottom Mid Right	105	X
Upper Mid Left	102	X	Upper Mid Right	100	X
Bottom Mid Left	103	X	Right Corner	4	X
Upper Center	101	Y	Bottom Mid Right	105	Y
Upper Mid Left	102	Y	Bottom Center	104	Y
Left Corner	5	Y			
Right Corner	4	Y			

Equation (6.2) was used to calculate the Smile Amplifier that we introduce in this study.

$$SmileAmplifier = \Sigma(fc_{s,a}[i][z] - B[i]) - \Sigma(fc_{s,s}[i][z] - B[i]) \qquad (6.2)$$

Where $fc_{s,a}$ holds the landmark points that are added for a smile, fcs,s holds the landmark points that are subtracted for a smile, i is the ID# of a given landmark point, and z is the current frame number, the Smile Amplifier performs a single-step calculation. The Smile Amplifier subtracts B[i], the baseline, from

the landmark point coordinates that are subtracted for a smile. This difference is summed for all of the points listed in the right half of Table 6.2. The result of this sum is subtracted from the sum of the differences between the points established for addition (the left half of Table 6.2) and their baselines.

To classify an expression as a smile, we compare the Smile Amplifier calculation to a set threshold. In reality, the Smile Amplifier gives a reading of the intensity of a smile. For the purposes of our work, we considered only its ability to detect the appearance of a smile by crossing a minimal threshold value.

Surprise

In classifying surprise expressions, we took a similar approach. Using the results of the tracking, we developed a list of landmark points on the face that increase above the baseline during a show of surprise, and we developed a list of points that decrease below the baseline during a surprise. These points are shown in Table 6.3 below. A total of 13 points on the eyebrows and on the outside of the mouth were used.

Equation (6.3) was used to calculate the Surprise Amplifier that we introduce in this study.

$$SurpriseAmplifier = \Sigma(fc_{srp,a}[i][z] - B[i]) - \Sigma(fc_{srp,s}[i][z] - B[i]) \qquad (6.3)$$

Where $fc_{srp,a}$ holds the landmark points that are added for a surprise, fcsrp,s holds the landmark points that are subtracted for a surprise, i is the ID# of a given landmark point, and z is the current frame number, the Surprise Amplifier performs a single-step calculation like that of the Smile Amplifier. The variables holding the landmark points are populated with values from Table 6.3.

Table 6.3: Coordinate addition/subtraction arrangement for surprises

Description	ID #	Axis	Description	ID #	Axis
Top Inner Lip	202	Y	Bottom Left Inner Lip	207	Y
Upper Mid Left	102	Y	Bottom Inner Lip	206	Y
Upper Center	101	Y	Right Corner Inner Lip	204	Y
Left Corner Inner Lip	200	X	Bottom Mid Right	105	Y
Left Point Right Brow	400	Y	Bottom Center	104	Y
Top of Right Brow	401	Y			
Right Point Right Brow	402	Y			
Top of Left Brow	301	Y			

To classify an expression as a surprise, we compare the Surprise Amplifier calculation to a set threshold. Like the Smile Amplifier, the Surprise Amplifier gives a reading of the intensity of the expression. For the purposes of our work, we considered only the ability to detect the appearance of a surprise by crossing a threshold value.

Anger

Similar to the previously described expressions, a list of points was developed for the determination of anger. These points are listed in Table 6.4.

A qualitative analysis of these points shows that they follow the expected pattern of lowered and drawn-in eyebrows for the expression of anger.

Table 6.4: Coordinate addition/subtraction arrangement for anger

Description	ID #	Axis	Description	ID #	Axis
Left Point Left Brow	300	X	Top of Left Brow	301	Y
Top of Left Brow	301	X	Right Point Left Brow	302	Y
Right Point Left Brow	302	X	Right Point Right Brow	402	X
			Top of Right Brow	401	X
			Top of Right Brow	401	Y
			Left Point Right Brow	400	Y

Disgust, Sadness, and Fear

Contrary to the conclusions drawn by the work of Ekman and Friesen [18], our preliminary work suggested that disgust and fear might not be universal expressions. Our attempts to determine common trends in the movement of tracked points proved difficult. Another complication arose when trying to induce these expressions. Subjects who were asked to express disgust or fear often had difficulty doing so on demand.

Our attempts to classify sadness were impaired by the tracking ability of FaceAPI. In most cases, a sad face involves the lowering of the corners of the mouth and the raising of the center of the lips. This configuration proved difficult to track in real-time.

6.2.3 Baselines and Classification Strength

The tracking experiments showed meaningful results. We were able to see consistent motion in several distinct points on all but 2 of the smiles that were tracked. We observed that the coordinates of points 5.x, 102.x, 103.x, 101.y, 102.y, 5.y, and

4.y increased in value during a smile, and the coordinates of points 105.x, 100.x, 4.x, 105.y, and 104.y decreased during a smile. These results allowed for the determination of accurate baselines and classifiers that were extended to a larger sample of subjects. As shown in the Figure 6.3, the Smile Amplifier, showed a nearly 10-fold increase in intensity over the signals from individual points.

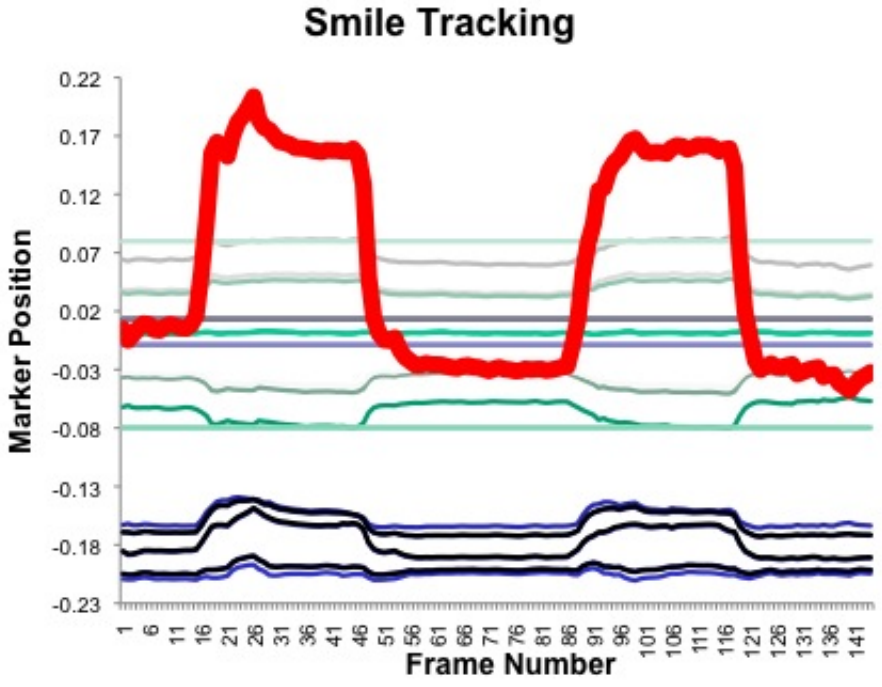


Figure 6.3: Results of smile tracking for one subject. The subject produced two smiles with about 1 second in between. The lighter lines track all the positions of the standard landmark points on the graph. The Smile Amplifier (bold red line) is superimposed on the graph, showing a nearly 10-fold increase in intensity over the individual signals.

We were interested in learning whether the Smile Amplifier could be used to detect poorly tracked faces. We retrieved a recorded series of poorly tracked

smiles from one subject, and applied the Smile Amplifier to the acquired signals. As shown in Figure 6.3, the Smile Amplifier was able to produce a weak (peak intensity = 0.063) yet observable signal from the data that originally did not show any detectable signs of a smile.

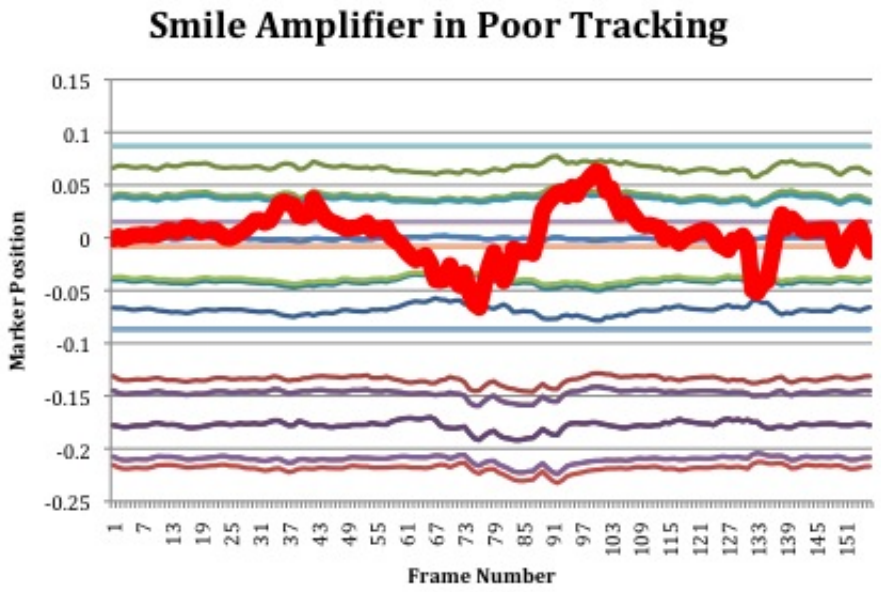


Figure 6.4: Results of smile tracking for a poorly tracked series of smiles. The subject produced two smiles with about 1 second in between, but the signals for a smile were not distinctly seen in the standard landmark points (the lighter lines on the graph). The Smile Amplifier (bold red line) is superimposed on the graph, showing a weak yet observable signal.

The 20 smiles of the initial 10 subjects were tracked with the Smile Amplifier. The baselines for the smiles were set at 0, and the start and end frame numbers were normalized between 1 and 2, respectively. As shown in Figure 6.4, all but two of the tracked smiles produced clear signals in the Smile Amplifier and exceeded 0.1 in value for the duration of the smile. 0.1 was selected as the value

for the systems detection threshold in this case.

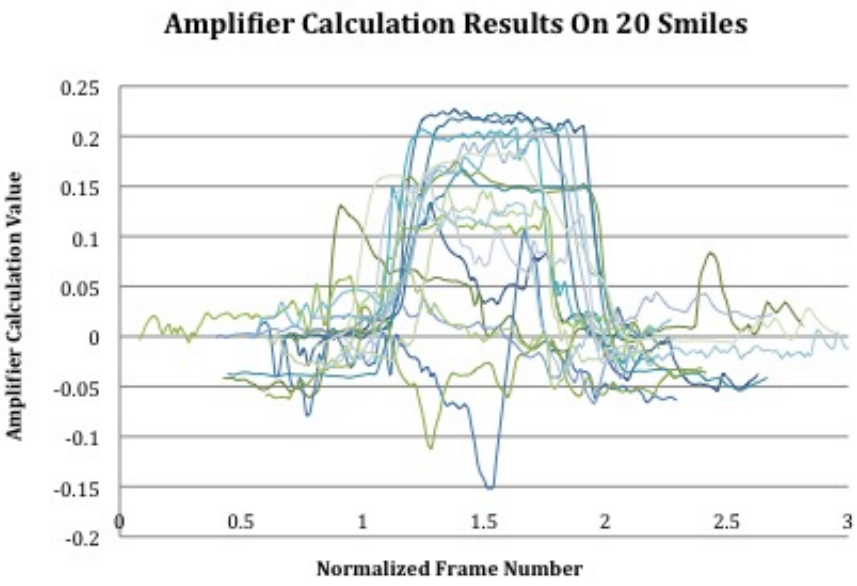


Figure 6.5: Smile Amplifier results for 20 distinct smiles. The baselines for the smiles are set at 0, and the start and end frame numbers are normalized between 1 and 2, respectively. All but two of the smiles showed significant Smile Amplifier values, exceeding the 0.1 mark.

A similar approach was taken for the tracking of surprised expressions. The results of the tracking for one subject are shown in the figure below. We were able to see consistent motion in several distinct points for the surprised expressions that were tracked. We observed that the coordinates of points 202.y, 102.y, 101.y, 200.x, 400.y, 401.y, 402.y, and 301.y increased in value during a show of surprise, and the coordinates of points 207.y, 206.y, 204.y, 105.y, and 104.y decreased during a show of surprise.

The Surprise Amplifier was calculated and superimposed on the raw data in Figure 6.6, showing a 4-fold increase in intensity over the largest of the individual

signals. This presented a clear and detectable signal that exceeded a threshold value of 0.1. The result was demonstrated to be repeatable with all 10 subjects.

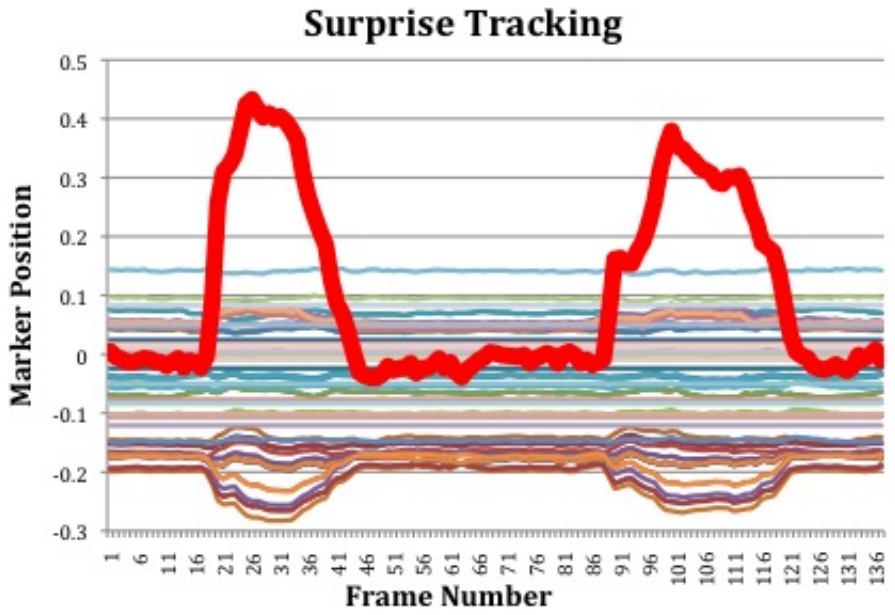


Figure 6.6: Results of surprise tracking for one subject. The subject produced two surprised expressions with about 1 second in between. The lighter lines track all the positions of the standard landmark points on the graph. The Surprise Amplifier (bold red line) is superimposed on the graph, showing a 4-fold increase in intensity over the largest of the individual signals.

Similar results were also seen for anger. However, as previously noted disgust, sadness, and fear were far less reliable.

6.2.4 Filtering

The real-time classification of expressions was unreliable and subject to influence by signal noise. In order to provide more reliable expression classification for users, a low-pass moving average noise filter was included.

For each frame, the previously described Expression Amplifiers were calculated. If the Anger Amplifier exceeded the established threshold, the expression for that frame would be set from Neutral to Angry. However, if the Surprise Amplifier exceeded its threshold, then the expression would be replaced with Surprise. Furthermore, if the Smile Amplifier exceeded its threshold, then the expression for that frame would be set to Smile. In this way, Smile is given priority over Surprise, which is given priority over Anger. If no Expression Amplifier exceeds its given threshold, then the expression for that frame remains at Neutral.

A running history of expressions over 30 frames (about 1 second at 30 frames per second) was maintained with this arrangement. In addition, a running history of activity/noise (represented by a sum of absolute values of deviations of all points from baseline) was maintained for the same duration. Each time that the current expression was read by the user interface, two operations were performed. First, the most common expression in the running history was found. Then, if the values in the activity history were within +/-30% of the first value of the activity history for more than 20 out of the 30 frames, the most common expression would be returned as the current expression. Otherwise, the current expression would be set to Neutral. This filtering provided the added advantage of allowing the analyzed person to talk (have lots of facial activity) without triggering rapid changes in the detected expression.

6.3 Graphical User Interface

We created a Graphical User Interface (GUI) to provide real-time metrics for debugging and improving the system. The GUI was written in the .NET user

interface application. By using the Windows Forms library in .NET, we created a Form and placed labels on it. The labels show the calculated Expression Amplifier values, the current expression, the currently recognized person, and head pose data among other metrics. Additional labels can be added to increase functionality.

6.3.1 System Interaction

Users are able to interact with the system through audio feedback and voice control from Microsofts Speech Recognition Engine and Microsofts Speech Synthesis Engine. System interaction is accomplished by using a Bluetooth headset as it has an in-ear speaker and a built-in microphone.

Audio feedback is mainly given in narrative form, but can also be toggled to use Microsoft system sounds for expression feedback. When a face is found, the system will alert the user by saying "face found." The system will also output the name of the person recognized when asked, or "no one" if the system does not find or recognize a face. Recognized commands will also be repeated back to user for confirmation of the recognition and execution of the command. If an expression is recognized, the system will output the expression to the user in either narrative or sound form, based on which feedback mode is set. In narrative form, "Smile" is outputted for happiness, "Surprised" is outputted for surprise, and "Angry" is outputted for anger. In sound form, an ascending sound (Sound On.wav) is played for happiness, an exciting sound (tada.wav) is played for surprise, and a descending sound (Windows Hardware Remove.wav) is played for anger. The default expression feedback setting is sound mode.

Moving on to how to control the system, the system recognizes commands

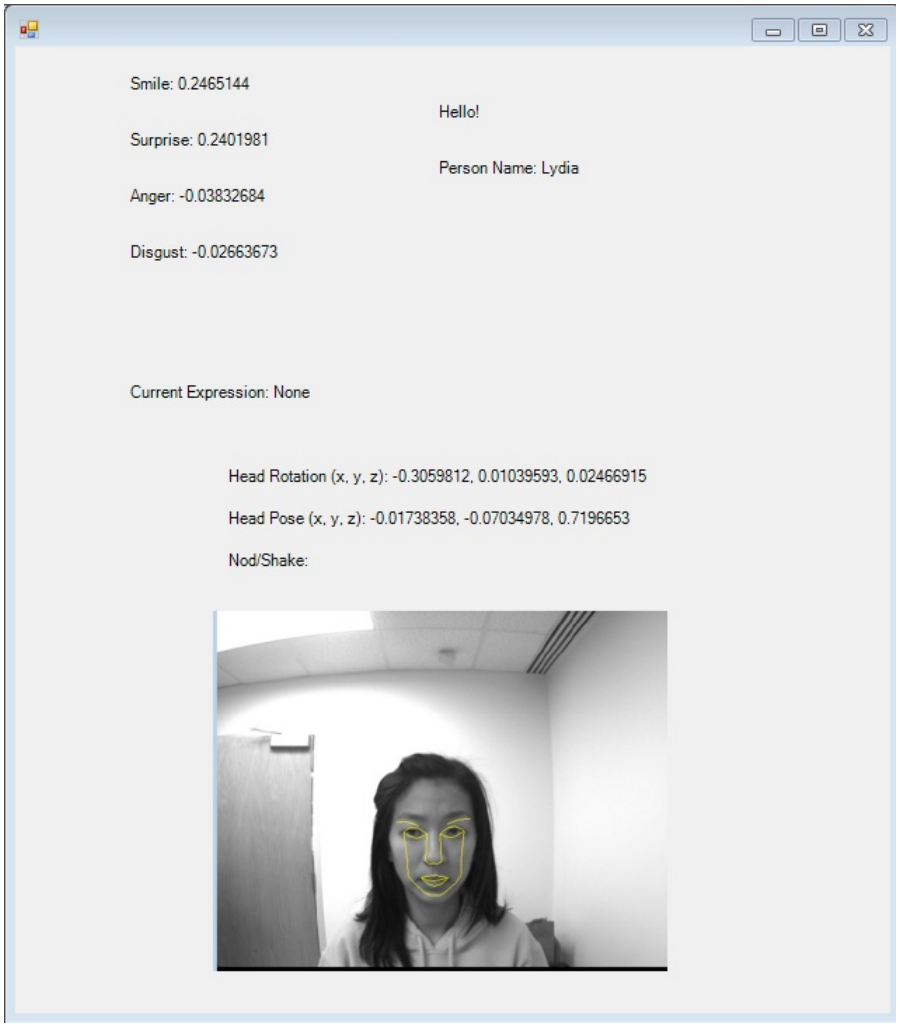


Figure 6.7: The interface is capable of indicating when a face has been found, who has been recognized, the expression data, and the rotation and pose of the head. These values are used to indicate expression shown and whether a shake or a nod of the head has been detected

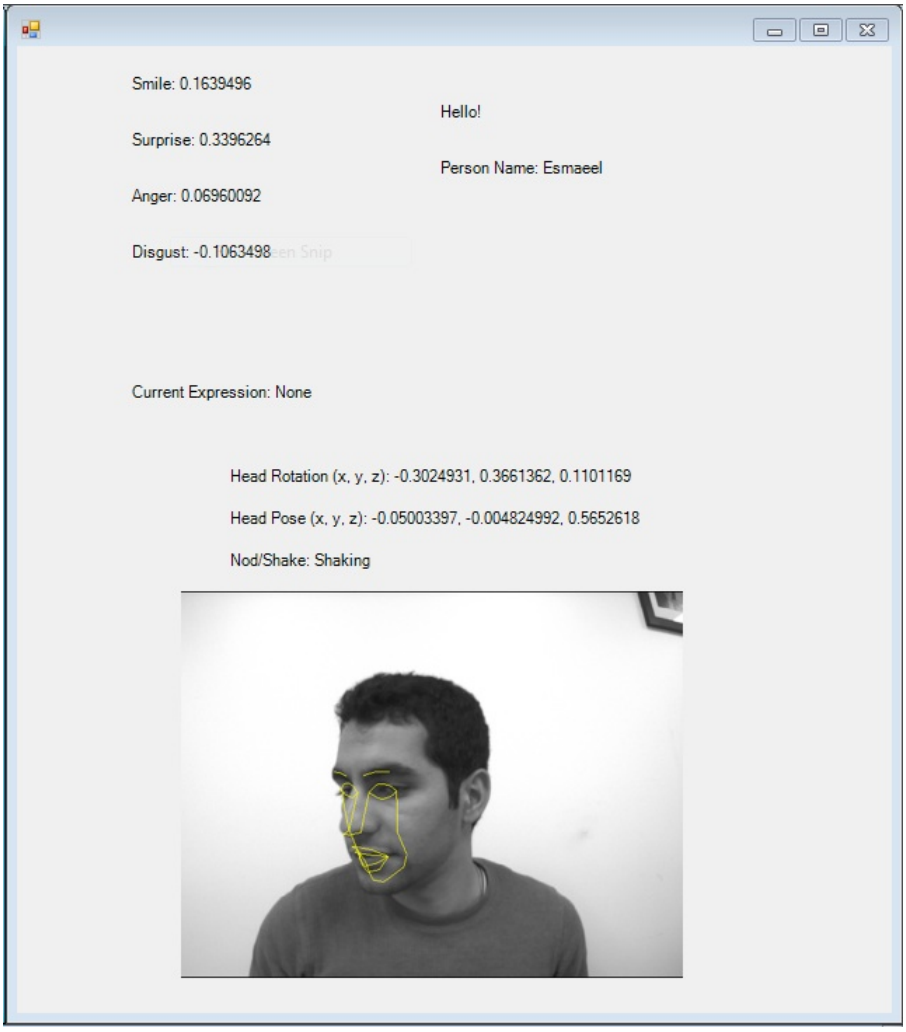


Figure 6.8: The interface is capable of identifying a face and maintaining tracked points as the head is rotated

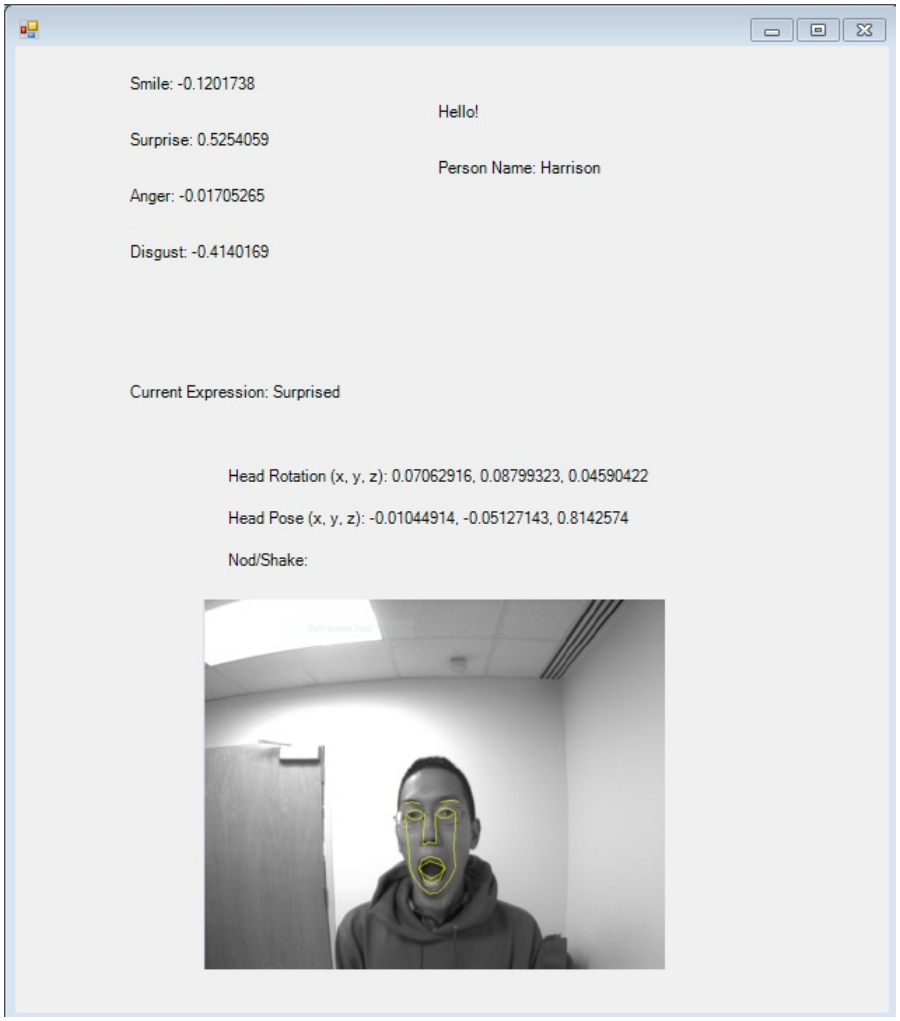


Figure 6.9: Here, the GUI indicates the person as well as the recognized surprise expression

that are part of enabled command sets called "grammars". Enabling different grammars at different times allows the system's speech recognition to be more accurate. The system currently has two grammars: a grammar for recognition commands and another grammar for changing the feedback mode.

To give the system a command, users must address the system as "computer," and then follow with one of the preset commands. As mentioned above, the system will repeat back to user the command it recognized, if any, to confirm the recognition and execution of a command. The two main preset commands include "Computer, who was recognized?", and "Computer, change feedback mode." The first command is used after the system has told the user "face found." The second command will turn off the recognition grammar and enable the feedback grammar, allowing the user to choose from either words (narrative) or chimes (sounds) for expression feedback. While the feedback grammar is enabled, users just have to say "change mode to words" or "change mode to chimes." After the system has notified the user that the feedback mode has been changed successfully, the feedback grammar will be disabled and the recognition grammar re-enabled. Finally, the command "Computer, stop listening" can be used to stop the application from running. This allows the user to turn off the system when it is not needed.

6.4 Accuracy Testing

6.4.1 Picture Database

A database of pictures was created in order to test the accuracy of our system in both recognition and expression situations. The database used during testing

Figure 6.10: Horizontal angles ranging from -30 to 30 degrees

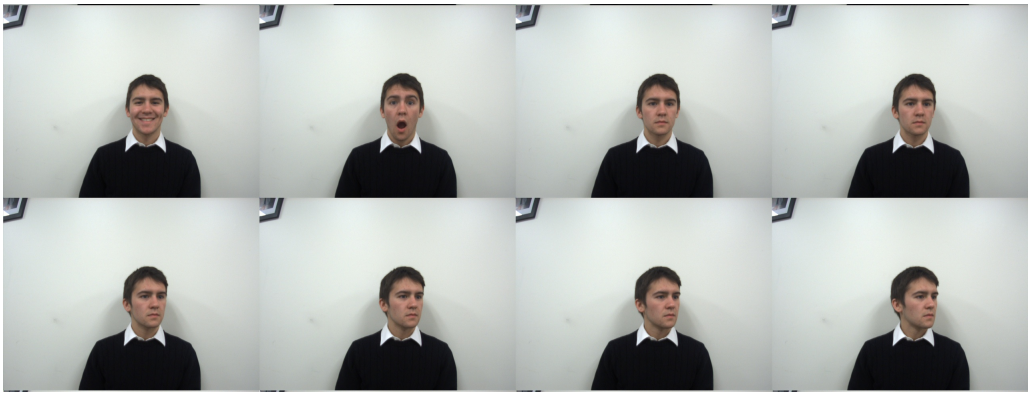


Figure 6.11: Expressions include a smile, two neutral pictures, and a surprised face

consisted of 260 total pictures, twenty pictures of all thirteen of our team members. The pictures for each person were all taken at a distance of three feet, and had a neutral expression except as noted. Pictures were taken at angles ranging from -30 to 30 in 5 degree intervals.

At an angle of 0, four pictures were taken two neutral pictures, one picture with the subject holding a surprised expression, and one picture with the subject holding a smile. Pictures were also taken at pitch angles ranging from -10 to 10 in 5 degree intervals. The total number of pictures for each person is then twenty.

Figure 6.12: Pitch angles ranging from -10 to 10 degrees

6.4.2 Accuracy Testing Methodology

In order to test the expression and recognition accuracy of the system, the following procedure was used. First, the test subject stepped into the frame of the camera at a distance of two feet and with a neutral expression. Then, the system was given twenty seconds to recognize the test subject. If the test subject was correctly identified for the duration of the twenty seconds, the trial was marked as a success. If the test subject was not identified in the twenty seconds, then the trial was marked as a fail. If at any point during the twenty seconds an incorrect person was identified, the trial was also marked as a fail.

After the first twenty seconds had concluded, the test subject was asked to hold a surprised expression for ten seconds. During these ten seconds, two variables were recorded: the person identified by the system, and the expression recognized by the system. If the correct person was identified for the duration of the ten seconds, the two feet recognition category was marked as successful. If, during the ten seconds, the wrong person was ever identified, the two feet recognition category was marked as a fail. If, for the duration of the ten seconds, the correct expression (surprise) was identified, the two feet expression category

was marked as a success. If, at any point in the ten seconds, an incorrect expression was recognized (anger or happiness), the two feet expression category was marked as a fail. After the ten seconds had expired, the test subject was asked to return their face to a neutral expression. The ten second test at two feet was then repeated two additional times for angry, and happy expressions.

After the test had been repeated for angry and happy expressions, the test subject was asked to remain looking at the camera with a neutral expression while they stepped back to a distance of four feet. The test subject was instructed to maintain their neutral expression and the system was again given twenty seconds to identify the test subject. If the correct person was identified for the duration of the twenty seconds, the trial was marked as a success. If an incorrect person was identified during the twenty seconds, the trial was marked as a fail.

After these twenty seconds had concluded, the test subject was asked to hold a surprised expression for ten seconds. During these ten seconds, two variables were recorded: the person identified by the system, and the expression recognized by the system. If the correct person was identified for the duration of the ten seconds, the four feet recognition category was marked as successful. If, during the ten seconds, the wrong person was ever identified, the four feet recognition category was marked as a fail. If, for the duration of the ten seconds, the correct expression (surprise) was identified, the four feet expression category was marked as a success. If, at any point in the ten seconds, an incorrect expression was recognized (anger or happiness), the four feet expression category was marked as a fail. After the ten seconds had expired, the test subject was asked to return their face to a neutral expression. The ten second test at four feet was then repeated two additional times for angry, and happy expressions. These tests

were conducted on every person included in the database, for a total of 182 trials 14 trials per person.

6.4.3 Accuracy Testing Results

Overall, the testing was considered a success. Facial recognition rates were successful 92% of the time, as shown in Table 6.5, and expression recognition rates would have also been successful 92% of the time if facial features had been tracked correctly in all tests. Unfortunately, many times facial features were poorly tracked during our trials, resulting in a much lower expression recognition success rate of only 65%. When the test subject was at the close range of two feet however, the expression recognition success rate, even when accounting for miss-tracked features, was a more acceptable 77%, as shown in Table 6.6

Table 6.5: Accuracy Testing for Recognition

Recognition Success - 2 Feet	Recognition Fail/Inconclusive - 2 Feet
96.15 %	3.85 %
Recognition Success - 4 Feet	Recognition Fail/Inconclusive - 4 Feet
88.46 %	11.54 %
Recognition Total Success	Recognition Total Fail/Inconclusive
92.31 %	7.69 %

6.4.4 Accuracy Testing Conclusions

At this point, we believe that recognition rates have reached a level of success that is acceptable for our application. Expression recognition rates however, still need to be improved upon before everyday use could be considered.

The main problem with expression recognition seems to be the tracking of the features of the face. As a team, we believe that the quality of the camera is not the limiting factor. We believe that the limiting factor is the facial feature tracking algorithms utilized in faceAPI. As the tracking algorithms become more sophisticated and accurate, the expression filters we designed should be accurate enough to permit everyday use for the blind community.

We would also like to clarify some discrepancies between our results as compared to the results of Xiao and Yang [17]. Xiao and Yang stated successful facial recognition rates of 99.2%, which is about 7% higher than our results. We believe that this difference can be accounted for with the fact that our system can not only calculate facial recognition, but can also calculate expressions, all in real time. We believe that if our system had been designed to only calculate facial recognition similarity, and had not been running in real time, we could reproduce similar success rates of 99%.

For future testing, we would like to conduct similar tests using a different number of pictures in the database for each trial. Twenty pictures of each person, at specific angles relative to the camera and with different expressions, is not feasible for a blind user to collect. Future tests should use an incrementally smaller number of pictures in the database in order to determine what the minimum number of pictures per person should be to maintain a high level of successful recognition rates. In order to improve expression recognition accuracy, we believe that faceAPI must either be abandoned or improved upon so that a more accurate tracking of facial features can be accomplished. With more accurate tracking of facial features, expression recognition should become accurate enough for blind users in real world situations.

Table 6.6: Accuracy Testing for Expressions

Surprise Success - 2 Feet	Bad Tracking	Bad Identification
92.31%	7.69%	0.00%
Surprise Success - 4 Feet	Bad Tracking	Bad Identification
76.92%	15.38%	7.69%
Surprise Total Success	Bad Tracking	Bad Identification
84.62%	11.54%	3.85%
Anger Success - 2 Feet	Bad Tracking	Bad Identification
84.62%	7.69%	7.69%
Anger Success - 4 Feet	Bad Tracking	Bad Identification
69.23%	30.77%	0.00%
Anger Total Success	Bad Tracking	Bad Identification
76.92%	19.23%	3.85%
Happiness Success - 2 Feet	Bad Tracking	Bad Identification
53.85%	38.46%	7.69%
Happiness Success - 4 Feet	Bad Tracking	Bad Identification
15.38%	61.54%	23.08%
Happiness Total Success	Bad Tracking	Bad Identification
34.62%	50.00%	30.77%
All Expressions Total Success	Bad Tracking	Bad Identification
65.38%	26.92%	7.69%
2 Feet Total Success	Bad Tracking	Bad Identification
76.92%	17.95%	5.13%
4 Feet Total Success	Bad Tracking	Bad Identification
53.85%	35.90%	10.26%

Chapter 7

Subject Testing the Prototype

In order to test the prototype and elicit feedback from potential users, five blind participants were recruited from various listservs and contacts. Four were female and one was male. All were above the age of 50. Some of participants had previously expressed interest in participating when they were interviewed last year. All were local and traveled to the University of Maryland campus, where the experiments were held.

Our primary objective in conducting experiments with blind individuals was to garner their reactions to our existing prototype, gain an understanding as to the best direction through which to develop the device, so that we could better fit our device to our target community. In conducting the experiments, there were further objectives examined. With our first experiment, we aimed to assess how well the blind participant is able to pick up facial expression cues, without any assistive technology. The second experiment was to evaluate how effective our feedback system is, using a 100% accuracy simulated system. The third experiment sequesters data regarding our facial recognition, while the final experiment gauges the comfort level with which the blind participant uses our actual device.

7.1 Methodology

The following subsections outline the methodology followed in assessing the effectiveness and usability of our device. Before the four experiments were conducted, participants were given 10 minutes to familiarize themselves with the device, 5 minutes with the expressions analysis system and 5 minutes with the recognition system. Throughout the experiments, team members explained how to use the device and allowed participants to test out the device on team members' faces and expressions. All participants signed consent forms digitally, and answered the interview questions below.

- How do you describe your social interactions with other visually impaired people? Sighted people?
- Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use? Rate how effective this device is from 1-5 (5 being the highest). Explain your rating.
- Using a scale of 1-5 (5 being the highest), rate how important it is for you to recognize emotional cues in your daily social interactions.
- Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize faces and convey identities to you. Explain.
- Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you. Explain.
- What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes,

etc.? What methods would pose significant inconveniences to you? Explain.

- On a scale of 1-5 (5 being the highest), how comfortable would you be carrying a cane mounted with a camera which connects to a laptop in a backpack in public? Explain.

- If you had such a device, would you wear it when interacting with other visually impaired people? With sighted people?

7.1.1 Experiment 1: Video Without Device

The purpose of this first experiment was to create a controlled experience in which participants collect nonverbal cues using their unassisted senses. Specifically, the first experiment was created to assess the participants ability to recognize expressions without any assistive technologies. In order to test this, participants listened to a video and were asked to use a clicker to identify each time they thought the speaker was smiling. Participants did this activity for two videos. The video clip was chosen such that the speakers face had to be shown for the majority of the clip and the facial expressions had to be clear and prominent. We searched different talk shows, and news interviews for such clips. The chosen videos were from an interview without the questions, so that the participants could focus only on the interviewee. Each clip lasted around 5 minutes.

We established the number of smiles seen in the video by having multiple team members watch the clips separately and identify the time points during which the speaker presents a smile. A smile was defined to occur when all three members unanimously agreed. The videos for experiment 1 were unaltered, but the video clips for experiment 2 were modified so that a specific sound (chime and

smile) was heard in sync with the predefined smiles. Participants were instructed to click a counter every time they believed the speaker in the video to be smiling.

7.1.2 Experiment 2: Programmed Cues

The purpose of this experiment was to evaluate the usefulness of two types of audio feedbacks for a 100% accurate expression analysis system (contrasted with the experience involving no assistive technologies). For the second experiment, participants simply listened to a second pair of video clips. These were the modified clips that played audio feedback synchronized with the pre-established smiles. Before beginning the experiment, participants were allowed to hear a sample of the audio feedback so that they knew what to expect when the video played. In our final design, there are multiple sounds that correspond to different expressions that can be detected. Due to time constraints during the experiments and the learning curve of associating particular sounds with specific expressions, participants were only tested with smiles. Following this stage was the first set of post-experimental questions. Participants were asked to rate/explain how obtrusive they found the device, what new information they gained from the device and how it may have changed the way they perceived the videos, how satisfied they were with the device and how they would modify it. The outline below describes the procedure of both experiments.

1. Explain this experiment to participant:

   Participant will watch a video (5 minutes). They will wear a bluetooth headset and every time the person in the video smiles or shows one of the 6 universal expressions, the system will tell the expression of the person in the video.

2. Play the video for the participant.

3. After video is over, ask participant following questions:

    - On a scale of 1-5 (5 being the highest), rate how obtrusive you found our feedback system.
    - Does the feedback give you new information that you couldn't have otherwise conceived?
    - If so, what new information did you gain and did it change the way you perceived the video?
    - On a scale of 1-5 (5 being the highest), rate how satisfied you are with our device.
    - How could we modify our device to better suit your needs?

7.1.3 Experiment 3: Recognition System Testing

This experiment was used to gauge blind participant feedback on the facial recognition system of prototype. In this experiment, participants were able to use the device to recognize team members. Three team members stood in front of the participant and the participant used the device to recognize each member. Then, the three team members randomly switched positions and the participant was instructed to use the device to find where each member was located again. Participants were then asked to respond on how useful they thought the recognition system was and how comfortable they were with using the system in public. The outline below describes the procedure of both experiments

1. Room setup: have a chair in the center of the room facing a row of three chairs (in positions labeled left, middle, and right).

2. Have the test subject sit in the lone chair in the center of the room and the 3 team members stand in front of the row of 3 chairs.

3. Have the 3 team members randomly switch positions.

4. New experiment conductor will give the following directions to the test subject: The three team members have now switched positions. Use the recognition device and tell me where each test conductor is seated.

5. After identifying all three team members positions, the team member will confirm or deny his/her position.

6. Have blind participant answer post-experiment questions and audiotape their responses:

    - For the following questions, please rate your opinion on a scale from 1-5 (5 being the highest):
        (a) This recognition system was useful.
        (b) This recognition system improved the ease of recognizing people.
        (c) I feel comfortable using this system in public.
        (d) I would use this system in a group setting.
    - What was the most useful part of our device?
    - What was the least useful part of our device?
    - How could we modify our device to better suit your needs?
    - Would you use this device in a real life setting? Why or why not?
    - Do you have any other feedback for us?

7.1.4 Experiment 4: Conversation with Expression System

This experiment assesses how comfortable the participants are with using the device in an open, one-on-one conversation. The participant held a brief, 2 minute, one-on-one conversation with a team member. During the conversation, the device was used to analyze the facial expressions of the team member and relay the expression feedback to the blind participant. Due to time constraints of the experiment and the learning curve of associating particular sounds with specific expressions, the device only relayed feedback for smiles. They were then asked to comment on how obtrusive they found the feedback system, what new information they gained that they couldnt have otherwise known, how useful and comfortable they were with the expression system and how satisfied overall they were with the device. The outline below describes the procedure of both experiments

1. Participant will be sitting in a chair. Team FACE member will be standing in front of another chair. Ensure that the participant and the team member are directly facing each other.

2. Explain the experiment to the participant.

3. Participant and Team FACE member will have a natural conversation. The camera is focused on the team member and is providing audio feedback about expression (smile only) to the participant.

4. After conversation, ask participants the following questions and audiotape their responses:

    - For the following question please rate your opinion on a scale from

1-5 (5 being the highest): Rate how obtrusive you found our feedback system.

- Does the feedback give you new information that you couldn't have otherwise known?
- If so, what new information did you gain?
- On a scale of 1-5 (5 being the highest), rate the following statements:
  (a) This expression system was useful.
  (b) This system improved my ability to recognize expression.
  (c) I feel comfortable using this system in public.
  (d) I would use this system in a group setting.
  (e) I was satisfied with the device.
- What was the most useful part of the expression system?
- What was the least useful part of the expression system?How could we modify our device to better suit your needs?

5. Ask if the participant has any further comments or questions.

7.2 Results

7.2.1 Participant 1

The participant in our first study was born blind with a condition known as retinopathy of prematurity. We showed the blind participant two brief videos where we had already identified the number of smiles present, and asked her to press the clicker each time she thought the speakers in the videos were smiling.

The first video she pressed the clicker one time; the second video she pressed the clicker four times. Our second experiment evaluated the effectiveness of our audio feedback system for expression analysis through another two videos. Much like the videos in experiment 1, the videos in experiment 2 also featured speakers who smiled several times throughout the recording. However, unlike those in experiment 1, the videos in experiment 2 were programmed with chime or voice feedback that corresponded to the presence of a smile. On a scale of 1-5, with 1 being the least obtrusive, the subject rated our chime feedback with a 1 and our voice feedback with a 2. She indicated that she had a slightly greater preference for the chime feedback, but acknowledged that it came with a steeper learning curve; ultimately, she said that she would be fine with either option. The subject ranked her satisfaction of our feedback system with a 5, 5 being the highest. She explained that the feedback allowed her to detect the presence of a face and identify the facial expression, both of which she would not have been able to accomplish in the absence of a conversation.

After experiment 2, we gave the participant time to familiarize herself with our prototype. She was able to use the bump markers that we placed on the cane to orient the camera in the direction she desired. For experiment 3, the participant was asked to use our system to identify three members of Team FACE. In order to recognize faces, users must give the system a series of verbal prompts, such as "initialize baseline" and asking it, "who was recognized?." The user can also enable the expression analysis by asking, "what is the current expression?" The participant had difficulty remembering the wording of the verbal prompts and the sequence in which the prompts needed to be said. The participant also found it hard to orient the cane in the direction of a face because the system did not

tell her when a face was detected; she moved the camera too fast over the three team members and, as a result, the camera and system did not have adequate time to detect and process the faces.

When asked if she could think of any improvements or modifications for our system, she suggested having chime feedback for expression analysis and voice feedback for facial recognition to better separate the two functions of our system. She also voiced concern about adjusting the camera to focus on a person taller or shorter than the height of the white cane that we used. In a meeting situation, we told her that she could unscrew the camera and place it on a tripod to focus on the faces of people seated at the same table. But in other scenarios, she wanted to see a prototype that was more versatile to accommodating different heights. She also noted that the weight of the camera on white cane could affect the mobility of blind users because of the redistribution of weight.

The participant wanted us to specify where our system should be used (e.g. quiet room, party, conference, etc.). She did not think our system would be feasible in a party setting because of the high noise level. Throughout all the experiments, the participant stressed the importance of making our system as discreet as possible, for her own safety and for her interactions with others. She did not want the system to affect her societal interactions, as she expected that others would be uncomfortable if they knew that the system was analyzing their faces. She said, "You want to be as natural and as though you were a sighted person. You don't want the blindness issue to get in there."

The subject said that it would be "cool to serve on jury" with this device and be able to analyze the expressions on the defendants or witnesss face. She also thought that our system had potential to be used in convicting attackers; if

a blind individual had this system and was attacked, then she thought that the blind individual could identify his or her alleged attacker through the system and give this information to police.

7.2.2 Participant 2

Our second participant was born blind and had a greater preference for our facial recognition feature. She stated multiple times throughout our entire experiment that she really wanted the facial recognition part of our system ("I want one of these. I've been wanting this for years.") During the first experiment, the participant clicked times for video 1 (which has smiles). She clicked times for video 2 (which has smiles). She found it hard to tell whether the speaker in video 1 was smiling, and she thought that the speaker in video 2 was smiling throughout the video because "he has a smiley voice."

In the second experiment, the subject found it surprising that individuals could smile in the middle of phrases, as opposed to at the beginning of a phrase like she had initially thought. She ranked the distraction of our chime feedback a 1 (the least distracting), saying that, "I'm used to things talking over each other. It doesn't bother me. I don't think you will find it to be obtrusive for a lot of blind people." She ranked the distraction of the voice feedback with a 2, indicating slightly more distraction than the chime feedback. She rated her satisfaction with our feedback system with a 4, with 5 being the highest. The participant explained that she thought "you could probably get enough info about people's experssions by their voice. I think the facial recognition (I'm anxious to see that), for a lot of blind people, would be the most important thing." When asked what improvements or modifications she would make to our system, she

suggested attaching the camera to something other than a cane (e.g. wrist band) for blind individuals who do not use a cane (for example, this participant uses a sighted guide dog).

When we told her that we wanted to see whether congenitally and non-congenitally blind individuals had a greater preference for one of the systems, she said, "I would think that maybe people who are blind later in life would find that they miss facial expressions. To me, it doesn't matter. A lot of times if I'm looking for a particular person and they are coming down the hallway towards me, I want to know before they are gone. I would prefer for facial recognition. Facial expression is interesting, but I don't know it would be something that I would lug in a backpack. If I could get facial recognition, I would carry a hell of a big device for that."

The participant suggested buying canes from Whiteplace and More or Ambutek and Revolutionary Enterprises because the canes from the NFB are typically taller than the ones many blind individuals use. During experiment 3, the subject suggested replacing voice recognition (saying "who was recognized") with a button that, when pressed, would prompt the recognition feature. The participant also recommended adding physical descriptions to the recognition system, such as descriptions of height and eye color. The only other modification she suggested was having the system tell her when the camera has captured a face.

When presented with multiple chime options, she said that she liked the happy chime feedback the best. She suggested putting our system on an iPhone; the participant stated that she would definitely be comfortable using our system in public, and she sees herself using it in her office to recognize people when they are coming (like sighted people can do) and to know before they leave

When asked for improvements that we can make to our feedback system, she recommended varying their frequency of feedback between 5 to 30 seconds instead of having a constant stream of feedback. She also suggested having a voice feedback that symbolizes the beginning and end of an expression ("begin expression" and "end expression"). The participant does not think the system would negatively impact her interaction with individuals. Rather, she thinks that the system will allow her to tailor the way she interacts with people based on their facial expressions; for example, if someone is happy, she would avoid talking about upsetting issues. We found the participant thinks more in terms of emotions rather than physical facial expressions; for example, she thinks in terms of whether somebody is happy as opposed to if somebody is smiling.

7.2.3 Participant 3

The subject of our third experiment was also born blind. At birth, he was able to see shadows and shapes but he slowly lost that ability between ten and twenty years of age. Thus, he never had the ability to recognize faces or facial expressions.

After introducing the basic pieces of our device, i.e. bluetooth, camera, cane, and laptop and explaining the basic premise of our overall experiment, the subject was asked to perform the first section of our experiment by clicking when he believed that the speaker in the selected video was smiling. Although there were actually number of smiles in the first clip, the subject believed that there were none by stating "the smile was not in the first one". Likewise in the second clip, the subject believed that there was only one smile when in fact, there were . It can be noted that the participant clicked when the interviewee in the clip was laughing, and thus the click was accurate. However whenever the interviewee in

the clip was just speaking, our participant was not able to distinguish any smiles.

Our next section of the experiment has videos with prerecorded smile indicators, either the spoken "smile" or a chime. Thus these videos were used to simulate that of a perfectly working prototype. The participant rated both systems as having a distraction of 2 (somewhat distracting). However, he believed that our chimes were actually more distracting than the spoken "smile". He explained that he prefers a much shorter chime, or "ding", instead of a sound with longer duration. Our participant also preferred a sound that is not heard often, so that the sound would not blend in with the environment.

When asked how satisfied the second participant was with the system, the participant responded with "when the system with indicating "smiles", it didn't seem that way to me ... it's not something that I would consider reliable". This was interesting in that a similar situation had arisen with our second subject. Thus we can propose a hypothesis that oftentimes when the feedback from device goes against a persons gut feeling (even if the feedback is accurate), opposition to the device can result. This occurred twice during our experiments, in which the participant believed that the expression feedback was "unreliable", since their gut feeling had told them a different story.

In the third section of the experiment, our blind subject was exposed to the facial-recognition portion of our prototype. Three team members stood in a line and the blind participant was told to point the camera at one of the members. The blind participant appeared to react significantly more positively to the facial-recognition portion of our prototype, repeating "yeah, I like this. I like this.", as the prototype started to recognize the members that the blind participant was pointing the camera to. When asked if the "recognition system improved the

ease of recognizing people", our blind participant responded with a 4 (improved significantly). In response to if the participant would feel comfortable using the system in public, the participant responded with a 5 (full comfort). He also indicated that he would feel fully comfortable using the system in a group setting. Regarding hardware, our third participant found that focusing the camera was the most useful part of the device. In comparison to the facial expression portion of our prototype, the facial recognition portion received more positive reviews from our blind participant. The participant reasoned that, "if I'm in a large group, I'd like to be drawn towards someone I know."

In the fourth section of the experiment, the blind participant was allowed to experience the functionalities of our current prototype, by having a general conversation with a team FACE member. The blind participant responded positively, rating that he would be highly comfortable using our current prototype in public (rating of 5, the highest rating). However, he still concluded that the facial recognition system is significantly more useful than the facial expression system, by rating the overall expression with an average rating of 3. Our third blind participant provided a rating of 4 for the usefulness of our overall system.

Our congenitally blind participant stated that he would prefer having facial recognition over facial expression, thus further confirming our hypothesis that we had generated from the blind interviews.

7.2.4 Participant 4

Our fourth participant found the feedback during videos to be obtrusive, and explained that she was unable to pay attention. Instead of the word "smile" he was more comfortable with hearing a chime associated with a smile being

shown. She found it "fascinating that the guy was smiling so much" and said the feedback provided new information to her. "Having just listened to the voice without knowing the smile part, I wouldn't have guessed that he was smiling so much. So it was helpful." The participant was satisfied with the feedback that, though distracting, was extremely helpful. She indicated a change in the chime would be better; a shorter chime is more effective and useful rather than a longer chime as was used in the experiments.

In response to questions about the facial recognition aspect of the device, the participant indicated it was useful, increase the ease of recognizing a person, and was comfortable to use in public and a group setting. The most useful part of our device for her was "seeing everyone run around." The participant was not sure if she would use such technology in general, not just during conversations. In response to questions on expression recognition, the participant reported that "knowing somebody is smiling or frowning is really good. I can only pick up sometimes from their voice or their way of talking."

In experiments with video clips, subjects did not learn anything new from audio feedback, "I'm not sure [the feedback gave me any new information], except that I disagreed with it. It didn't sound to me as if the person being interviewed was smiling. It sounded completely and totally serious." Feedback did not change the way subject perceived video. As such, this participant was different than all other experiments as she distrusted the technology. The subject found the feedback to be "puzzling" because she did not think the speaker was smiling. She did not numerically rate satisfaction with device, but said it was "unreliable" because it did not match what she heard in speaker's voice. Later, upon receiving explanation, subject revealed that she rated our system as "unreliable" because

she thought, "smiles means happy." Her explanation was that "[she] couldn't detect, from my own gut, happiness there. And that's why I responded the way I did."

7.2.5 Participant 5

The subject of our fifth experiment was born blind, and thus was not ever able to see facial expressions or recognize faces. She appeared to find that facial recognition and facial expression recognition have equal importance commenting that both are integral to her life.

During the first subsection of our experiment, the subject was shown two separate clips and was asked to click when he believed that the speaker in the selected clip was smiling. The subject commented that "the person sounds so serious" in the first clip, despite the fact that there were several smiles. When shown the second video, the subject only clicked once during when the interviewee in the video was smiling, despite the fact that the interviewee had actually smiled times.

According to our subject, "Just from the tone of voice, it just sounds like the person is very serious". Thus, it seems as if the subject really focused on fluctuations in the voice to tell when the interviewee was smiling. There were several times in the video though, where the fluctuation was not as apparent, and thus the smile was missed by our subject.

The subject was then shown videos using first chimes, then the word "smile", to indicate actual smiles in the video, to simulate a 100% accurate working prototype. When asked if the feedback was obtrusive, the subject rated it a 1, claiming that the simulated feedback was "not obtrusive" at all. The subject also found

that she had gaineed new information from the feedback system, describing it as giving her "more information as far as what the facial expression was". Our subject was highly satisfied with out feedback system, providing a 5 on satisfaction. Our subject suggested that we should incorporate different types of sounds for different facial expressions, rather than chimes.

The blind subject was exposed to the facial-recognition portion of our prototype in section 3 of our experiment. She also ranked the importance of the facial-recognition portion a 5, indicating that "it would be really important." The subject was also very satisfied with our current prototype and found that the chimes were not obtrusive at all, with a rating of 5. The subject was also very comfortable with using the system in public, giving the comfort level a rating of 5. She commented that "[she] would use it [and] wouldn't find any problem with it." The main concern that she had was that our prototype might be a little too bulky for everyday use and she suggested that we should try to design a "handheld device." She also found no problem with using the device in a group setting. The subject ranked our overall prototype with a 4.

7.2.6 Analysis

For some blind people, the term "smile" may not be distinct from "happiness." With the aid of vision, sighted individuals are aware that the physical expression on one's face does not always match his/her emotional state (perceived through his/her voice/tone). Given the same directions as the other blind participants, subject 3 instinctively assumed that smiling was synonymous with happiness. This suggests that blind individuals are in tune with facial expressions in differing degrees. In such case, this would ultimately present as another factor for the

Table 7.1: User Preferences for Audio Feedback

Subject	Intrusion (1-5)	Satisfaction (1-5)	Impressions
1	1-2	No comment	Could glean facial expression without relying solely on conversational context
2	1-2	4	Did not know that people could smile while talking
3	2	-	Was confused that the speaker could smile without necessarily feeling happiness
4	4	3	Would not have guessed that the speaker was smiling as frequently
5	1	5	No comment

learning curve that comes with our device. However, we still cannot conclusively eliminate miscommunication between subject and experimenter as the cause of this incident. To address both possibilities, we should, henceforth, strive to be more clear to blind users about the abilities, and especially, limitations of our device.

Just as blind people are specific about their preferences in cane features (as learned from our second meeting with the NFB), they are sensitive to the chimes and noises that they are exposed to. To better improve the usability of our device,

as suggested by subject 3 (and possibly others), we should offer a variety of brief and distinct chime selections.

7.3 National Federation of the Blind

We again partnered with the National Federation of the Blind to obtain feedback on the developed system in early December 2011. Five team members attended the meeting to discuss our product with 3 NFB representatives. We brought our entire system to the meeting the camera attached to the end of the cane with both facial recognition and expression recognition capabilities. The NFB representatives were extremely impressed with the product, especially the facial recognition portion of the software. The team members were able to take pictures of the NFB representatives, store their pictures in the database, and immediately perform facial recognition real time. The demonstration was able to perform at 100% accuracy. It performed so well that the representatives requested the NFB president to stop by and test our product. When the camera recognized the presidents face, he commented that this was the only product that he has seen that actually works! Our expression recognition software was still in the beginning developmental stages, so that portion of the demonstration did not perform as well as we hoped. However, the representatives understood that we were still in our development and testing phase and were excited for us to return with a completely working product. They also provided commentary on the design of our hardware. Though two of the three representatives did not mind the cane being attached on the end of the cane, the third representative preferred a detachable camera. He expressed how difficult or awkward it would be to walk around with a weight on the end of his cane, and having to manuver a cable that is attached

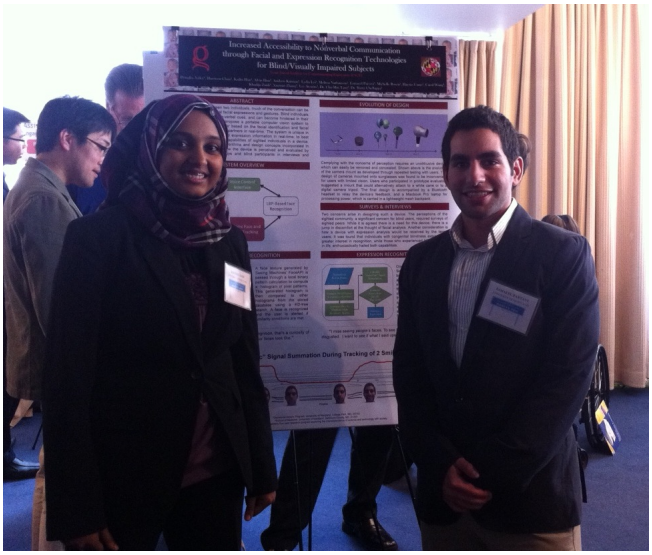
between his cane and a backpack. Though he was against the design, the other two representatives did not mind the extra weight and stated that the weight and cable did not disturb their balance.

Following the meeting, the NFB representatives extended us an invitation to attend their Research and Development (R and D) day in February 2012, to provide a demonstration of our device to a team of engineers. Our affiliation with the NFB has also contributed to a similar invitation from the American Association of Visually Impaired Lawyers in July 2012.

Chapter 8

Publications

Our first paper was submitted to The 13th International ACM SIGACCESS Conference on Computers and Accessibility, held in Dundee, Scotland, in October 2011.



Two members traveled to the conference, which explores information and assistive technologies for elderly people and those with disabilities. The paper was written as a proposal to present a poster at the conference relating our software development. The poster focuses on the algorithms and factors used to analyze and output expressions that are recognized by the system. The poster

also discussed the surveys, interviews, and proposed subject tests. The ACM paper and associated poster are related in the following pages.

In the near future, we aim to submit literature based on the combined development of the software and hardware of our assistive technology device to the Journal of Visual Impairment and Blindness (JVIB). JVIB is an interdisciplinary journal which records information and issues faced by the blind and visually impaired community. The publication is issued monthly with research articles and news coverage. Our hope is that this will garner interest in science and technology and the application of assistive technologies.

Increased Accessibility to Nonverbal Communication through Facial and Expression Recognition Technologies for Blind/Visually Impaired Subjects

Douglas Astler [*] Harrison Chau Kailin Hsu

Alvin Hua Andrew Kannan Lydia Lei

Melissa Nathanson Esmaeel Paryavi Michelle Rosen

Hayato Unno Carol Wang Khadija Zaidi

Xuemin Zhang Cha-Min Tang [†]

ABSTRACT

Conversation between two individuals requires verbal dialogue; the majority of human communication however consists of non-verbal cues such as gestures and facial expressions. Blind individuals are thus hindered in their interaction capabilities. To address this, we are building a computer vision system with facial recognition and expression algorithms to relay nonverbal messages to a blind user. The device will communicate the identities and facial expressions of communication partners in realtime. In order to ensure that this device will be useful to the blind community, we conducted surveys and interviews and we are working with subjects to test prototypes of the device. This paper describes the algorithms and design concepts incorporated in this device, and it provides a commentary on early survey and interview results. A corresponding poster with demonstration stills is exhibited at this conference.

Categories and Subject Descriptors

A.1 [**Introductory and Survey**]: Miscellaneous; D.2.10 [**Software Engineering**]: Design

General Terms

Algorithms, design, human factors

[*]Equal authorship to Team F.A.C.E. of the Gemstone Honors Program, University of Maryland, College Park. Team may be contacted at face.gemstone@gmail.com

[†]University of Maryland School of Medicine, Baltimore

ASSETS'11, October 24–26, 2011, Dundee, Scotland, UK.
ACM 978-1-4503-0919-6/11/10.

Keywords

Assistive technologies, computer vision, blindness, face recognition, expression recognition

1. INTRODUCTION

Humans rely on visual cues when interacting within a social context, and for those who cannot see, their lack of sight interferes with the quality of their social interactions [3]. Since the majority of non-verbal communication is comprised of facial gestures and expressions, blind individuals are at a disadvantage when engaging in interpersonal communication [3].

Since over 80% of human communication is actually carried out nonverbally [1], there is an opportunity to include image processing in assistive technologies. These missed messages include facial expressions, and hand gestures, which cannot be detected, and more importantly interpreted, without sight. With computer vision, an assistive device used by blind individuals can include algorithms designed to "read" faces. The design of our prototype includes realtime feedback for facial recognition in addition to algorithms built for identification of six different facial expressions.

2. ALGORITHMS

2.1 Facial Recognition

In the design of our device, fast recognition time is an important factor. A standard recognition system, under controlled conditions, can achieve a recognition rate of 99.2% in near realtime [5]. One system, developed by Krishna et al., uses a Principal Component Analysis (PCA) algorithm for facial recognition [3]. Our system uses a commercial solution (FaceSDK; Luxand, Inc., Alexandria, VA) for face recognition, matching images to user-built databases, and can provide audio or haptic feedback in realtime. We are pursuing interconnectivity with a user's facebook account

to automatically populate the image database with the profile pictures of friends.

2.2 Expression Analysis

One classic expression detector is a parametric optical flow-based algorithm developed by Yacoob and Black in 1997. This computationally intensive algorithm tracks facial features in a sequence of images to detect the presence of six expressions [4]. These expressions are described according to the six universal expressions theory - happiness, anger, sadness, disgust, surprise and fear - as developed by Ekman and Friesen. [2]

Many traditional expression detectors, including Yacoob and Black's, have been too slow to be used for realtime detection. In an effort to increase usefulness, our device has been designed to overcome the previous performance limitations. At the hardware level, a high speed IEEE-1394b camera is used to capture images at a very high frame rate while minimizing CPU load. We use commercial software (FaceAPI; Seeing Machines Inc., Canberra, Australia) to accurately track facial features in realtime. By analyzing the movement of facial features, we were able to find common trends in the same expressions among different individuals. We created computationally undemanding thresholding algorithms to classify expressions based on the detected trends. Using this method we have had success in detecting expressions at near 30 frames per second for happiness (Figure 1b), surprise, disgust, and anger.

3. INTERVIEWS AND SURVEYS

In addition to the hardware and software considerations, two societal concerns arise in the design of this device. The direct needs of the intended users must be considered, and that includes the perceptions of the sighted community, which may also be a concern for blind users. Our survey of over 200 sighted individuals showed agreement that there is a need for this device, but there is a jump in discomfort at the thought of facial analysis. Receptivity by the blind community was evaluated by interviewing nine subjects. We found that those who experienced blindness later in life expressed a greater interest in emotion recognition, while individuals with congenital blindness enthusiastically hailed both capabilities. One blind individual was quoted saying, "I lost my vision 14 years ago. I miss seeing people's faces. To see if they are angry, happy, or disgusted." Interviews also suggested features – such as mounting the system's camera atop a white cane (Figure 1a) – desired by the blind community.

4. DISCUSSION/FUTURE DIRECTIONS

Using data collected in surveys and interviews, we hope to tailor this device to the specific needs of all who will come in contact with it. In the next several months, we are working to have blind users test a prototype in a controlled environment. These experiments will allow us to evaluate the system's accuracy and usefulness. Whereas an assistive device to recognize people was met with predicted enthusiasm, one that analyzes facial features raises concerns of privacy and misinterpretations. Another issue is the effect on a user's standard behaviors while using the device. Combined, these factors could lead to an overall negative effect on communication. It is these elements that increase the challenge in developing an unobtrusive device as a social aid for the blind. Our goal is to develop an effective system to provide a blind user with the same analysis of faces that sighted individuals benefit from. As one sighted individual surveyed wrote, "The application of the [device] would be similar to another person looking at me and analyzing my facial expressions with his or her sight." While we cannot restore vision, we seek to balance communication between sighted and blind individuals, to enhance the quality of blind individuals' societal interactions, and to provide a device that fulfills all needs without interfering with comfort.

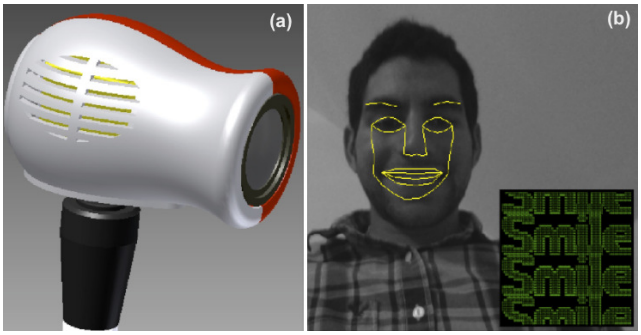


Figure 1: (a) Our design of a camera atop a standard white cane. (b) The expression recognition identifying a smile.

5. ACKNOWLEDGMENTS

Thanks is due to Team F.A.C.E.'s mentors and advisors Dr. R. Chellappa and Dr. C.M. Tang, and graduate advisor Mr. L. Stearns. We also thank the Gemstone Honors Program at the University of Maryland. Finally, we would like to thank the National Federation of the Blind for their support.

APPENDIX

A. ADDITIONAL AUTHORS

B. REFERENCES

[1] Argyle, M. Salter, U. Nicholson, H. Williams, M. Burgess, P. The communication of inferior and superior attitudes by verbal and non-verbal signals. *British Journal of Social Psychology*, 9:222–231, 1970.

[2] Friesen, W.V. Ekman, P. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2):124–129, 1971.

[3] Little, G. Black, J. Panchanathan, S. Krishna, S. A wearable face recognition system for individuals with visual impairments. In *7th Annual ACM Conference on Assistive Technologies*, pages 106–113, 2005.

[4] Yacoob, Y. Black, M.J. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, 25(1):23–48, 1997.

[5] Yang, X.D. Xiao, Q. A facial presence monitoring system for information security. *IEEE Workshop on Computational Intelligence in Biometrics*, 2009.

Increased Accessibility to Nonverbal Communication through Facial and Expression Recognition Technologies for Blind/Visually Impaired Subjects

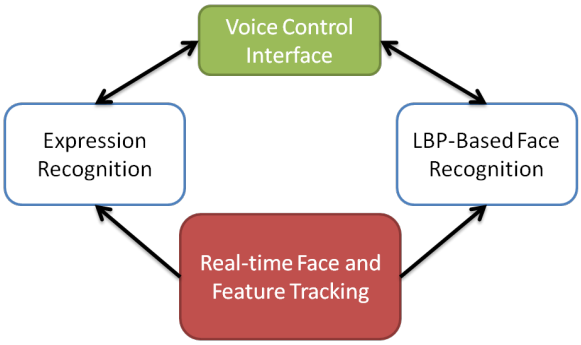
Team Facial Analysis for Communicating Expression (FACE)

Douglas Astler[1], Harrison Chau[1], Kailin Hsu[1], Alvin Hua[1], Andrew Kannan[1], Lydia Lei[1], Melissa Nathanson[1], Esmaeel Paryavi[1], Michelle Rosen[1], Hayato Unno[1], Carol Wang[1], Khadija Zaidi[1], Xuemin Zhang[1], Lee Stearns[1], Dr. Cha-Min Tang[2], Dr. Rama Chellappa[1]

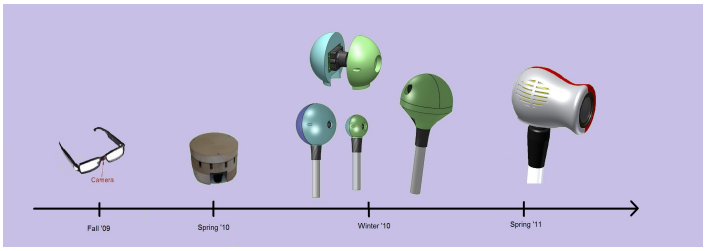
ABSTRACT

In verbal dialogue between two individuals, much of the conversation can be relayed through non-verbal facial expressions and gestures. Blind individuals do not have access to non-verbal cues, and can become hindered in their interaction capabilities. We propose a portable computer vision system to relay feedback to a blind user based on the facial identification and facial expression of communication partners in real-time. The system is unique in providing both recognition and expression information in real-time, to best emulate the communication capabilities of sighted individuals in a device. This poster describes the algorithms and design concepts incorporated in this device, in addition to how the device is perceived and evaluated by sighted participants in surveys and blind participants in interviews and subject testing of the device.

SYSTEM OVERVIEW



EVOLUTION OF DESIGN

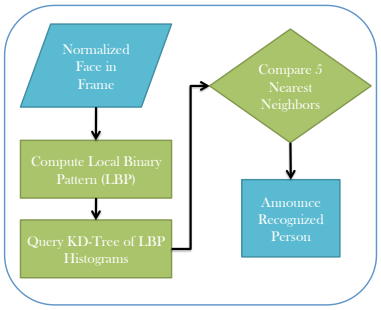


Complying with the concerns of perception requires an unobtrusive device which can easily be removed and concealed. Shown above is the evolution of the camera mount as developed through repeated testing with users. The design of cameras mounted onto sunglasses was found to be inconvenient for users with limited vision. Users who participated in prototype evaluation suggested a mount that could alternatively attach to a white cane or to a digital camera tripod. The final design is accompanied by a Bluetooth headset to relay the device's feedback, and a Macbook Pro laptop for processing power, which is carried in a lightweight mesh backpack.

SURVEYS & INTERVIEWS

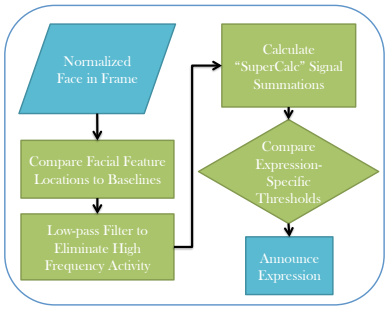
Two concerns arise in designing such a device. The perceptions of the sighted community, a significant concern for blind users, required surveys of sighted peers. While it is agreed there is a need for this device, there is a jump in discomfort at the thought of facial analysis. Another consideration is how a device with expression analysis would be received by the target users. It was found that individuals with congenital blindness expressed a greater interest in recognition, while those who experienced blindness later in life, enthusiastically hailed both capabilities.

LBP-BASED FACIAL RECOGNITION



A face texture generated by Seeing Machines' FaceAPI is passed through a local binary pattern calculation to compute a histogram of pixel patterns. This generated histogram is then compared to other histograms from the stored database using a KD-tree search. A face is recognized and the user is alerted if similarity conditions are met.

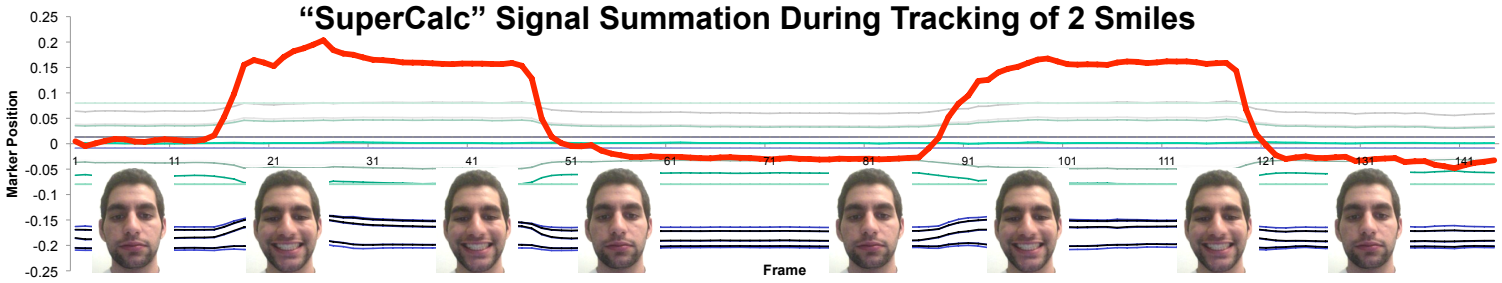
EXPRESSION RECOGNITION



Once a face is found through Seeing Machines' FaceAPI, a baseline is established. Points are then tracked in real-time for each frame and compared to the baseline conditions. If a change spans enough frames, then the "SuperCalc" signal summations of the points are compared to a series of established thresholds, and the facial expression is determined.

"People like me who've seen…for facial recognition, that's a curiosity of totally blind people, to know what your faces look like."

"I miss seeing people's faces. To see if they are angry, happy, or disgusted. I want to see if what I said upset them or made them happy."



[1]Gemstone Honors Program, University of Maryland, College Park, MD, 20742
[2]School of Medicine, University of Maryland, Baltimore County, MD, 21250

Gemstone is a multidisciplinary four-year research program exploring the interdependence of science and technology with society.

Chapter 9

Conclusions

While our device is functional, useful, and robust, there are still some improvements that could be made in the future. In our product design research, one of the most requested features of our device was reduced size. With the current state of mobile computing, processors are not fast enough to perform the algorithms utilized by our device in real time. But in the future, several advances could be made to enable this.

As technology progresses and mobile processors become more powerful, our device could potentially be driven by a smaller device. Another possibility would be to harness the power of cloud computing. With emerging 4G technologies such as LTE and WiMAX, the speed and reliability of wireless communication is increasing. Current mobile devices are capable of capturing video in a high resolution. If we can send these high resolution videos to a server across wireless networks, we can push computation onto the cloud, enabling our device to be delivered in a smaller form factor. Unfortunately, at this time, fast wireless networks are only just emerging, but this could be a viable possibility for our device in the very near future.

Another improvement suggested to us by blind users was the addition of a

button for user input. Currently, for our device, a user must ask the device, Computer, who was recognized? using their voice. In certain situations, this would be socially unacceptable. Instead, we could add a button (or multiple buttons) to our device to trigger certain commands.

Finally, for an assistive device to be adopted by a community, it cannot serve only one function. While our device does supply face recognition and expression analysis information, if it could provide extra information, blind users would be much more likely to use it. Previous groups have created assistive devices for navigation and object recognition for blind users. Ideally, these devices could all be integrated with ours, creating a single device with multiple assistive capabilities.

Our prototype will allow the visually impaired to greatly improve their social interactions with both sighted and blind people. The prototype gives the visually impaired facial recognition and expression information through a variety of means, verbal cues and chimes. Along with providing the users with additional information for personal use, it also provides them with information about their safety. No other device provides this level of freedom in assessing nonverbal cues to the visually impaired. No other device provides the real-time feedback and control over how this information is related.

The project cannot give sight back to the visually impaired but we can give them information about faces and expressions. This project puts the visually impaired one step closer towards equality with sighted people in conversation. New technologies will allow people with disabilities to improve their standard of living and put them closer to parity with people without disabilities. Here science, technology, and society intersect providing a better life for the visually impaired.

Chapter 10

Internal Review Board Documentation

10.1 Sighted Survey Consent Form

Project Title

Facial Analysis for Communicating Expressions

Why is this research being done?

This is a research project being conducted by Gemstone Team Facial Analysis for Communicating Expressions (FACE) at the University of Maryland, College Park. We are inviting you to participate in this research project because you are above 18 years of age and do not experience any form of blindness. The purpose of this research project is to make a portable camera and computer system that will allow a blind individual to recognize faces he/she knows, as well as detect whether the person being viewed is happy, sad, angry, surprised, fearful, or disgusted. We are seeking suggestions from possible users of our device so that we can build an effective device which fits all needs.

What will I be asked to do?

You are asked to fill out a brief online survey. You will be asked to express your opinion on this project. The purpose of this survey is to understand the

perception of the blind by the sighted community.

What about confidentiality?

We will do our best to keep your personal information confidential. To help protect your confidentiality, we will (1) keep all computer data entries on password-protected computers, (2) keep all files in a locked team office, (3) refer to participants by a number in our data entries and files, and (4) only allow Gemstone Team FACE mentors and members any access to data.

If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.

In accordance with legal requirements and/or professional standards, we will disclose to the appropriate individuals and/or authorities information that comes to our attention concerning neglect or potential harm to you or others.

What are the risks of this research?

There are no known risks associated with participating in this research project.

What are the benefits of this research?

The most tangible benefit of this research is the development of a wearable device to detect faces and expressions. We hope that, in the future, other people may benefit from our research through improved understanding of computer vision and how it can be used to aid communications between blind persons, as well as communication between blind and sighted persons.

Do I have to be in this research?

May I stop participating at any time?

Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

What if I have questions?

This research is being conducted by Dr. Rama Chellappa at the University of Maryland, College Park. If you have any questions about the research study itself, please contact Dr. Chellappa at:

If you have questions about your rights as a research subject or wish to report a research-related injury, please contact: Institutional Review Board Office, University of Maryland, College Park, Maryland, 20742; (e-mail) irb@umd.edu; (telephone) 301-405-0678

This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.

Statement of Age of Subject and Consent Your completion of the survey indicates that:

you are at least 18 years of age;,

the research has been explained to you;

your questions have been fully answered; and

you freely and voluntarily choose to participate in this research project.

Do you agree to the above terms? You must agree to proceed to the survey questions.

1. Yes, I agree

2. No, I do not wish to participate in this survey

10.2 Sighted Survey Questions

1. Age

2. Gender

   (a) Male

   (b) Female

3. Household Income

   (a) $0-$19,000

   (b) $20,000-$39,000

   (c) $40,000-$59,000

   (d) $60,000-$79,000

   (e) $80,000-$99,000

   (f) $100,000-$149,000

   (g) $150,000+

4. Race (Select all that apply)

   (a) American Indian/Alaskan Native

   (b) Black or African American

   (c) Asian or Pacific Islander

   (d) Hispanic/Latino

(e) White

(f) Other or Not Reported

5. Education (Highest degree obtained)

   (a) Some High School

   (b) High School Diploma/GED

   (c) Some College

   (d) Associate's

   (e) Bachelor's

   (f) Master's

   (g) Doctorate/Professional School

   (h) Other

6. How often do you interact with the blind?

   (a) Never

   (b) Rarely

   (c) Sometimes

   (d) Often

7. Do you think there are any societal stigmas associated with the blind? Explain.

8. How would you react if you saw a person carrying a cane mounted with a camera, which connects to a laptop in a backpack?

   (a) Uncomfortable

(b) Initially Uncomfortable

(c) Neutral

(d) Intrigued

9. Please explain.

10. How do you think you would react if you knew that such a device was analyzing your facial expressions and identifying your face?

(a) Uncomfortable

(b) Initially Uncomfortable

(c) Neutral

(d) Intrigued

11. Please explain.

12. Would such a device negatively affect any interactions you may have with the blind? Explain.

10.3 Blind Interview and Experiment Consent Form

University of Maryland College Park

 Initials_____ Date_____

Project Title	*Facial Analysis for Communicating Expression*
Purpose of the Study	*This research is being conducted by Dr. Rama Chellappa, Dr. Cha-Min Tang, and the members of the Gemstone Team Facial Analysis for Communicating Expression (FACE) at the University of Maryland, College Park. We are inviting you to participate in this research project because you are above 18 years of age, experience blindness, and stand to benefit from this study. The purpose of this research project is to make a portable camera and computer system that will allow a blind individual to recognize faces he/she knows, as well as detect whether the person being viewed is happy, sad, angry, surprised, fearful, or disgusted. We are seeking suggestions from possible users of our device so that we can build an effective device which fits all needs..*
Procedures	*The procedures involve participating in an individual interview or focus group. An interview will be a one-on-one discussion session with one participant and one researcher, while a focus group will be a discussion setting with five participants and two researchers. Questions in both settings encompass an assessment of existing social interactions between the blind and blind and/or blind and sighted in society, the effectiveness of current assistive technologies in social interaction, and potential design ideas for the device. Prior to conducting an interview or focus group, participants will be asked for information regarding their age, gender, socioeconomic status, nationality, and level of education to provide researchers with background information.* *The researchers will conduct interviews and focus groups as is convenient for the participants. The interviews will take approximately fifteen to twenty minutes to complete, while the focus groups will last approximately thirty to forty minutes.* *This research project involves making audiotapes of you to record responses for the researchers' reference. Only Gemstone team mentors and members will have access to these tapes and, upon completion of our project, these tapes will be destroyed.* *The researchers will also be asking participants to experiment with the device under controlled conditions. Participants will be asked to use the device while watching a video of an individual(s) speaking, as well as when conversing with a Team FACE member. The prototype will track facial features and convey expression cues to participants. Participants will then be asked to answer post-experiment questions to gauge the effectiveness of our device. Additionally, they will be asked to use the device to identify*

	individuals on Team FACE in a spatial context. Another series of post-experiment questions pertaining to the effectiveness of the device will follow. *___ I agree to be audiotaped during my participation in this study.* *___ I do not agree to be audiotaped during my participation in this study.*
Potential Risks and Discomforts	*There are no known risks.*
Potential Benefits	*The benefits to you include the development of a portable device to detect faces and expressions. We hope that, in the future, other people might benefit from this study through improved understanding of computer vision and how it can be used to aid communications between blind persons, as well as communication between blind and sighted persons.*
Confidentiality	*Any potential loss of confidentiality will be minimized by (1) keeping all computer data entries on password-protected computers, (2) keeping all files in a locked team office, (3) referring to participants by a number in our data entries and files, and (4) only allowing Gemstone Team FACE mentors and members any access to data.* *If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.*
Medical Treatment	*The University of Maryland does not provide any medical, hospitalization or other insurance for participants in this research study, nor will the University of Maryland provide any medical treatment or compensation for any injury sustained as a result of participation in this research study, except as required by law.*
Right to Withdraw and Questions	*Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.* *If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the investigator, Rama Chellappa at: Rm. 4411 A.V. Williams, 115 Paint Branch Drive, University of Maryland, College Park, MD, 20742; (301) 405 – 3656; rama@cfar.umd.edu and face.gemstone@gmail.com*
Participant Rights	*If you have questions about your rights as a research participant or wish to*

	report a research-related injury, please contact: University of Maryland College Park Institutional Review Board Office 1204 Marie Mount College Park, Maryland, 20742 E-mail: irb@umd.edu Telephone: 301-405-0678 *This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.*	
Statement of Consent	*Your signature indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study. You will receive a copy of this signed consent form.* *If you agree to participate, please sign your name below.*	
Signature and Date	PARTICIPANT NAME [Please Print]	
	PARTICIPANT SIGNATURE	
	DATE	

10.4 Blind Interview Questions

10.4.1 Demographic Questions

1. Age
2. Gender
    (a) Male
    (b) Female
3. Household Income
    (a) $0-$19,000
    (b) $20,000-$39,000
    (c) $40,000-$59,000
    (d) $60,000-$79,000
    (e) $80,000-$99,000
    (f) $100,000+
    (g) Prefer not to answer
4. Race (Select all that apply)
    (a) American Indian/Alaskan Native
    (b) Black or African American
    (c) Asian or Pacific Islander
    (d) Hispanic/Latino
    (e) White

(f) Other or Not Reported

5. Education (Highest degree obtained)

   (a) High School Diploma/GED

   (b) Associate degree

   (c) Bachelor's degree

   (d) Master's degree

   (e) Doctoral degree

   (f) Other

10.4.2 Interview Methodology

1. Distribute Braille consent form and ask blind subjects to carefully read it on their own. Direct them to sign it if they give consent and collect consent form upon completion

2. Assign participant a number and ask for participant's name

3. Record name and assigned number in Team FACE notebook

4. Write participant's assigned number on Braille demographic questionnaire and give to participant to complete

5. Upon completion, collect questionnaire

6. Inform participant that the interview will begin and that you will be tape recording the interview

7. State participant's assigned number into tape recorder

8. Ask the following interview questions and continue recording answers on a tape recorder; allow participant to elaborate whenever they feel inclined to (ask follow-up questions as necessary):

   (a) How do you describe your social interactions with other blind people? Sighted people?

   (b) Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use? Rate how effective this device is from 1-5 (5 being the highest). Explain your rating.

   (c) Using a scale of 1-5 (5 being the highest), rate how important it is for you to recognize emotional cues in your daily social interactions.

   (d) Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize faces and convey identities to you. Explain.

   (e) Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you. Explain.

   (f) What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.? What methods would pose significant inconveniences to you? Explain.

   (g) On a scale of 1-5 (5 being the highest), how comfortable would you be wearing sunglasses embedded with cameras which connect to a laptop in a backpack in public? Explain.

   (h) If you had such a device, would you wear it when interacting with other blind people? With sighted people?

9. Thank the participant for their cooperation and inform them that a Braille transcript of his/her interview will be available for pick-up at the Columbia Lighthouse within two weeks

10. Ask participant to provide feedback on the transcript once they receive it

11. Set date and time for pick-up with this participant

12. Personally deliver the Braille transcript to participant on assigned date and time at the Columbia Lighthouse and ask participant to provide feedback-/comments to the contact information that we have provided along with the transcript

10.5 Blind Experiment Methodology

10.5.1 Pre-Experiment

1. Participants are given 10 minutes to familiarize themselves with the device, 5 minutes with the expressions analysis system and 5 minutes with the recognition system.

2. FACE team members will explain how to use the device and participants can test the device on the team members i.e. Participants can ask team member to show different emotions to see how the device responds to the different expressions

3. Prior to the experiments, email participants consent form and have them send back a signed copy upon their consent (digital signature will count as their consent)

4. Assign participants numbers to maintain confidentiality

5. If participants have not been interviewed yet, ask the following questions and audiotape their answers:

    - How do you describe your social interactions with other visually impaired people? Sighted people?
    - Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use? Rate how effective this device is from 1-5 (5 being the highest). Explain your rating.
    - Using a scale of 1-5 (5 being the highest), rate how important it is for you to recognize emotional cues in your daily social interactions.
    - Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize faces and convey identities to you. Explain.
    - Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you. Explain.
    - What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.? What methods would pose significant inconveniences to you? Explain.
    - On a scale of 1-5 (5 being the highest), how comfortable would you be carrying a cane mounted with a camera which connects to a laptop in a backpack in public? Explain.
    - If you had such a device, would you wear it when interacting with other visually impaired people? With sighted people?

6. Show participants our prototype and explain how it works

7. Allow participants approximately 15 minutes to acquaint themselves with the prototype (assistance can still be given during this time)

10.5.2 Experiment 1: Clicker without Device

1. Have participants sit in front of a screen

2. Inform participants that a video with individual(s) speaking will be played (video will last around 5 minutes)

3. Instruct participants to click a counter each time they believe the speaker in the video smiles

4. Play video and allow participants to perform the task

10.5.3 Experiment 2: Clicker with Programmed Cues

1. Explain this experiment to participant:

   Participant will watch a video (5 minutes). They will wear a bluetooth headset and every time the person in the video smiles or shows one of the 6 universal expressions, the system will tell the expression of the person in the video.

2. Play the video for the participant.

3. After video is over, ask participant following questions:

   - On a scale of 1-5 (5 being the highest), rate how obtrusive you found our feedback system.

- Does the feedback give you new information that you couldn't have otherwise conceived?
- If so, what new information did you gain and did it change the way you perceived the video?
- On a scale of 1-5 (5 being the highest), rate how satisfied you are with our device.
- How could we modify our device to better suit your needs?

10.5.4 Experiment 3: Recognition System Testing

1. Room setup: have a chair in the center of the room facing a row of three chairs (in positions labeled left, middle, and right).
2. Have the test subject sit in the lone chair in the center of the room and the 3 team members stand in front of the row of 3 chairs.
3. Have the 3 team members randomly switch positions.
4. New experiment conductor will give the following directions to the test subject: The three team members have now switched positions. Use the recognition device and tell me where each test conductor is seated.
5. After identifying all three team members' positions, the team member will confirm or deny his/her position.
6. Have blind participant answer post-experiment questions and audiotape their responses:
    - For the following questions, please rate your opinion on a scale from 1-5 (5 being the highest):

(a) This recognition system was useful.

(b) This recognition system improved the ease of recognizing people.

(c) I feel comfortable using this system in public.

(d) I would use this system in a group setting.

- What was the most useful part of our device?
- What was the least useful part of our device?
- How could we modify our device to better suit your needs?
- Would you use this device in a real life setting? Why or why not?
- Do you have any other feedback for us?

10.5.5 Experiment 4: One-on-one Conversation with Expression System

1. Participant will be sitting in a chair. Team FACE member will be standing in front of another chair. Ensure that the participant and the team member are directly facing each other.

2. Explain the experiment to the participant.

3. Participant and Team FACE member will have a natural conversation. The camera is focused on the team member and is providing audio feedback about expression (smile only) to the participant.

4. After conversation, ask participants the following questions and audiotape their responses:

- For the following question please rate your opinion on a scale from 1-5 (5 being the highest): Rate how obtrusive you found our feedback system.
- Does the feedback give you new information that you couldn't have otherwise known?
- If so, what new information did you gain?
- On a scale of 1-5 (5 being the highest), rate the following statements:
  (a) This expression system was useful.
  (b) This system improved my ability to recognize expression.
  (c) I feel comfortable using this system in public.
  (d) I would use this system in a group setting.
  (e) I was satisfied with the device.
- What was the most useful part of the expression system?
- What was the least useful part of the expression system?How could we modify our device to better suit your needs?

5. Ask if the participant has any further comments or questions.

Chapter 11

Participant Transcripts

11.1 Interview Transcripts

11.1.1 Subject 1

Team FACE Member: So I have a few quick questions that are probably the easier ones that we can begin with. You don't have to answer any of these if you don't want to. They're just a demographic questionnaire. Can I ask your age?

Subject 1: 55.

Team FACE Member: Ok. And your gender?

Subject 1: Female.

Team FACe Member: And level of income. You can also not answer this if you're ok with that. We have from A through F. There are choices: A) 0 19,999 B) 20,000 39,999; C) 40,000- 59,999; D) 60,000 79,999; E) 80,000 99,999; and F) over 100,000 a year

Subject 1: You mean household income or just mine?

Team FACE Member: I'd say household income.

Subject 1: It's the last one-F.

Team FACE Member: Ok Great. Ethnicity? We also have A through E

in terms of choices: A) American Indian/ Alaskan Native; B) Black or African American; C) Asian or Pacific Islander; D) Hispanic/ Latino; E) White; F) Other or Not reported

Subject 1: White-E.

Team FACe Member: And level of education?

Subject 1: Four years of college.

Team FACE Member: So bachelor's degree?

Subject 1: Yes.

Team FACe Member: Ok, great. Thank you very much. And now we can move on to the other questions. First question I have here is how do you describe your social interactions with other visually impaired people?

Subject 1: Is this free form or do you give me choices?

Team FACE Member: Free form.

Subject 1: I'm pretty comfortable with other visually impaired people. I enjoy interacting with visually impaired people.

Team FACE Member: Do you ever think there's anything missing?

Subject 1: No. I think a lot of times things are missing when dealing with sighted people more than with blind people.

Team FACE Member: Right that was the second part of the question I was going to ask you. So how do you describe your interactions with sighted people?

Subject 1: They're ok if it's someone I know. But people that I don't know, well you miss the eye contact and they're uncomfortable with the lack of eye contact. Like in meetings and dealing with strangers. They're sort of uncomfortable sometimes so that's an issue.

Team FACE Member: Let's move on to question B. Do you currently use any

assistive technologies or devices that help you in your social interactions?

Subject 1: Social interactions? I mean with the computer sure. But that doesn't really help with my social interactions. Not quite sure what you mean.

Team FACe Member: We were just wondering because the device that we're planning on developing is going to be used for primarily social interactions. We're looking to develop something that'll analyze facial expressions.

Subject 1: No I don't have anything like that now.

Team FACE Member: Ok, we don't have any follow up questions for that. Let's move on to the next question. Using a scale of 1-5, 5 being the highest, rate how important it is for you to recognize emotional cues in your daily social interactions.

Subject 1: Since I've never had the ability to do it, you know visually. It really doesn't matter all that much. I'd rather have facial recognition. When people come up to you, you only have the split second when they say, "'Hi (Subject 1),"' to find out who they are. Whereas they've got, "'Oh, so-and-so is walking up to me. That's so-and-so."' But I don't have that ability. Being able to recognize faces and interacting them and know who they were based on just that real quick "'Hi (Subject 1)"' without them telling me who they are. That would be more important to me.

Team FACE Member: Rate your interest, on a scale of 1-5, 5 being the highest, in a portable device that could recognize faces and convey identities to you.

Subject 1: Yes! That would be 5. That would be really cool. Then you could interact with them. You would know who they were and maybe they'd be somebody you'd want to talk to. And when they're 15 feet down the hall from you then, "'Oh, that was so-and-so. Darn it!"'

Team FACE Member: Ok. Well, thank you for that. There is one more question which I feel is a little redundant. Rate your interest, on a scale of 1-5, 5 being the highest, in a portable device that could recognize and analyze facial expressions and convey these cues to you.

Subject 1: 2. I mean, you could tell a lot of that from somebody's voice.

Team FACE Member: I'm sorry. That first question where we asked you to rate how important it is for you to recognize emotional cues in your daily interactions, did you want to give a number for that as well?

Subject 1: 3, I guess because it's important to do it by someone's voice but I would give it a 2 for a machine that would do it.

Team FACE Member: Well the question is, just in general, how important it is for you to recognize these cues?

Subject 1: I would say a 3.

Team FACE Member: What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.?

Subject 1: They're so many. You could have it speak. You could have any number of things. You could have it vibrate. In a quiet room you wouldn't want it to say, "'Someone is smiling,"' if you're in the middle of a meeting or something. You could either have it be able to turn on and off. Just like a phone. You could have it either ring or vibrate. That would be fine. Then you would have to have tones for the vibrations.

Team FACE Member: Are there any methods that would pose significant inconveniences to you?

Subject 1: Well certainly, if it was really noisy. If it only had an audible

component and you couldn't turn it off. If it was really big and you had to carry it around. That would be inconvenient.

Team FACE Member: And what's your feeling on having it, for example, chime instead of straight up speaking to you?

Subject 1: But a chime is only a one thing. There wouldn't be a lot of deviation in different kinds of facial features. For speech, you have it say, "'Someone is smiling, someone is frowning, someone is crying, someone is laughing,"' with a chime you only have one thing unless you have all sorts of different kinds of chimes. And that would be harder to remember. There would be more of a learning curve for the chime.

Team FACE Member: On a scale of 1-5, 5 being the highest, how comfortable would you be carrying a cane mounted with a camera connecting to a laptop in a backpack in public?

Subject 1: I wouldn't mind but I don't want to carry all that stuff. And I don't use a cane anyway. I use a guide dog. I mean it could be something that you clip somewhere with belt clip. Like a camera that you clip to your pocketbook strap or such. Like I have a GPS receiver clipped to my portable computer thing that I carry with me. I wouldn't be really comfortable carrying something big like a laptop, no. I mean my shoulders are shot as it is. Not that big.

Team FACE Member: Would you give this a 1 out of a 1-5?

Subject 1: Yeah, if it was going to be that big I wouldn't do it.

Team FACE Member: What if you took the laptop out of it, and it all ran off of your existing cell phone.

Subject 1: That would be fine because I carry my cellphone with me all the time anyway. And that would be a 4. I would do that

Team FACE Member: And we have the last question If you had such a device, would you wear it when interacting with other visually impaired people? This is for the original question, so if there was a cane with a mounted camera connected to a laptop in a backpack.

Subject 1: I wouldn't carry something that big with a laptop. I probably wouldn't use it in interactions with visually impaired people because we tend not to use or pay attention to facial things as much. You know we would talk about those things. I would use it more with sighted people.

Team FACE Member: Ok.

Subject 1: And if it was small. I wouldn't ever carry a laptop with this device.

Team FACE Member: Thank you for your cooperation.

11.1.2 Subject 2

Team FACE Member: How do you describe your social interactions with other visually impaired people?

Subject 2: My social interactions are pretty mellow in general. Easy going.

Team FACE Member: This is with sighted people as well?

Subject 2: That's right.

Team FACE Member: Anything else?

Subject 2: Open minded. Very social. I'm very persuasive sometimes. I just ask for respect and equal opportunity.

Team FACE Member: Are your interactions generally on the good side? Or the great side? Is there anything missing when you interact with either other visually impaired individuals or sighted individuals?

Subject 2: It's neutral with both groups. It's not really negative or it's not

always extreme. There are good times and negative times. I basically think that's just life.

Team FACE Member: Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use? Rate how effective this device is from 1-5, 5 being the highest.

Subject 2: My cane. I use a note taker. It has voice-speech. I use my talk and screen reader.

Team FACE Member: How would you use your cane in social interactions?

Subject 2: It helps me to get around. To find obstacles in my way. To find steps, like if you're in the subway up on the platform. Or outside, you know, with curves. Helps me to be mobile so I can come and go like they can. My note taker, I have my daily planner on it. It has my medical information, it has a phone book so I can take calls on lunch breaks or emergencies. It just helps me to keep life organized.

Team FACE Member: For your note taker, how would you rate how effective this device is from 1-5, 5 being the highest?

Subject 2: 5, I'm lost without that thing.

Team FACE Member: Do you use any other devices?

Subject 2: I use a talk and screen reader at work, home and school. Screen reader reads everything on the computer screen. It enables the totally blind to navigate the computer just like a sighted person, just without looking at the screen. We can do internet research, Microsoft Word, word processing, spreadsheets. It just enables us to be independent and be able to navigate a computer.

Team FACE Member: And how would you rate this device, from 1-5?

Subject 2: 5.

Team FACE Member: Using a scale of 1-5, 5 being the highest, rate how important it is for you to recognize emotional cues in your daily social interactions.

Subject 2: 5. It could be a safety issue.

Team FACE Member: Emotional cues. For example, you could pick up on person's tone. Like if someone's excited, their voice might be a little higher pitched. Something that tells you about somebody's emotions.

Subject 2: 4 and a half. Like say if they're angry and you can pick up on that, you'll know not to say anything or not to ask too much of them. Or you'll just give them their space. Because if you try to interact with someone who's having a hard time, you might get a rude awakening. Well, you might end up upsetting them more. You don't want to do that. That's the last thing you want to do. Or you might be able to help pick them up. Who knows? I guess that's important, that we be able to do that. To pick up on their cues. If a stranger's angry, you don't want to mess with them, you don't know what's in their life, you don't know where they're at in their life. But if it's somebody you know, say like your child is upset about something, you want to be able to know that so you could help calm your child down and help get them back on the right path.

Team FACE Member: Rate your interest, on a scale of 1-5, 5 being the highest, in a portable device that could recognize faces and convey identities to you.

Subject 2: That would be a 5. I would be very excited if there were technology like that. They already have devices now that can identify money, not the coins because we can feel the coins and tell which one, but the bills. like say you had a fifty dollar bill, you slide it in this little device and hit a button and it'll tell you, "'fifty dollars."' That's very important. We need to know how much money we have on us. That helps us keep track of our budget. We have to be able

to survive and do all the normal things like the regular sighted people do. We just have to do it through these devices. Without these devices, we really can't survive. Because we would just be in some institution behind closed doors with the lights turned out on us. No, life is supposed to be enhanced to the maximum. And if that means we have to have technology to help us do that then that's what it means. There's no reason why anyone should be behind closed doors just because of who they are or what disability they have.

Team FACE Member: Rate your interest, again, on a scale of 1-5, in a portable device that could recognize and analyze facial expressions and convey these cues to you.

Subject 2: That would be 5.

Team FACE Member: Any explanation?

Subject 2: Like I said, it's very important. Especially if we don't have the visual acuity to really see the facial expressions, it's very important that we be able to hear. That's our strongest sense. And we have in addition the other three that help, that become more keen in the absence of that one sense. So in order for us to be able to comprehend things, make good decisions and live life to the fullest, it's important if we can't see it, we should be able to feel it or hear it or experience it through a different outlet just like someone's sight.

Team FACE Member: Our idea is that sometimes people could be very emotional but remain silent.

Subject 2: You wouldn't know. And you'd go ahead and say, "'Hello, how are you doing?"' or you ask them a question, and say they're in a very depressed state or they're pissed. They'll howl off. We got to be able to pick up on these cues. It could be a matter of safety. Safety should be number one in everyone's

case.

Team FACE Member: What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc? What methods would pose significant inconveniences to you?

Subject 2: Significant conveniences or inconveniences?

Team FACE Member: Both.

Subject 2: I'd like to have it audio announced. Say like if they have hands, and where we can feel the different vibrations and tones. A deaf and blind wouldn't be able to hear it but they could feel it through those tones and vibrations. With a combination of tones, motion, auditory.

Team FACE Member: But for you personally?

Subject 2: I'd like to have it that way too. That way I would understand. The motion along with the auditory, it would make sense to me. Most people nowadays have more than one way of learning things.

Team FACE Member: So with face recognition, it'll be pretty straightforward. You'll want to know that person's'name. It'll probably be audio. It'll say, "'Bob"' to indicate that Bob is five meters away. On the other hand, with expression, there are so many expressions that a person could convey. So how would you want this device to give you facial information?

Subject 2: Ok. Let's say a person is in a positive mode, a positive mood, then a positive expression will have a higher pitched voice, will have a higher beep when it's describing it to you. And if it's lower pitched, I'll know that the person is having a hard time in some way.

Team FACE Member: We were thinking about chimes because sometimes noises can be interfering, so we want something that's short and straightforward.

But what about vibrations for expressions? What do you think? For example, you could wear a belt with vibrators attached to them. Maybe a vibration on the right side could mean somebody is happy and a vibration on the left side could mean that somebody is in a negative mood. Because sometimes you don't want sounds like chimes going off. This is something that we want to ask you about.

Subject 2: Yeah, like if you're going to a meeting. You got a crowd of people. You don't want them to hear audio things calling everybody's name, that'll freak people and cause more emotions. Or maybe the belt could have a Braille display too for those blind people who read Braille. I would like a series of buttons where I can turn the speech off, put it in Braille mode. The option to turn Braille mode on, or speech mode on, or speech and Braille or different beeps. Or we can put an ear piece on. Not every totally blind person reads Braille. Some just do a lot of auditory.

Team FACE Member: So a device with multiple options?

Subject 2: Yes.

Team FACE Member: On a scale of 1-5, how comfortable would you be carrying a cane mounted with a camera connecting to a laptop in a backpack in public?

Subject 2: And I'd be wearing the backpack?

Team FACE Member: Right. So how would you feel about wearing and carrying this kind of device around in public?

Subject 2: Kind of awkward at first. But as I got used to it I wouldn't worry about it. I mean if it's to enhance my life, I can't worry about what people think. It's not about them all the time. Anything new is going to feel awkward until you become used to it and become fluent in it and then its not going to be a big

deal anymore.

Team FACE Member: At that point, you won't mind?

Subject 2: No.

Team FACE Member: If the benefits outweigh the negatives?

Subject 2: Right.

Team FACE Member: How about if the backpack was taken out of the equation and this was running off your own cell phone which you could slide into your pocket?

Subject 2: I would feel even more comfortable with that. That way I wouldn't have so much stuff to carry around.

Team FACE Member: Can you rank these? Both scenarios with the 1 through 5 scale.

Subject 2: Ok, the device connected to the cane and the backpack I'd give that about a three. And the device that's small like a cell phone I'd give it a 5. The smaller the device, the more powerful the device. Everybody's going micro now.

Team FACE Member: If you had such a device, would you wear it when interacting with other visually impaired people? With sighted people?

Subject 2: I'd wear it all the time. I'd show them this new piece of technology. Whenever I learn something new I want to put the information out there. Because the more aware, the more educated, the more we can make better decisions and interact more appropriately with each other. We shouldn't be looking at the world like half blind, half sighted. This is a world we all live in. We need to get along and respect each other. And survive together.

Team FACE Member: Thank you. This concludes the interview.

11.1.3 Subject 3

Team FACE Member: How would you describe your social interactions with other visually impaired people?

Subject 3: No, I hardly see other blind people. Except a blind couple who are close friends or at places that employ blind people.

Team FACE Member: How would you describe your social interactions with sighted people?

Subject 3: I see a whole lot of them. I have a whole family that is sighted. I interact with people at my church. I use Metro Access a lot. I interact with the drivers and other clients.

Team FACE Member: Do you currently use any assistive technology that help you with social interactions?

Subject 3: I use email and Internet.

Team FACE Member: How about face-to-face?

Subject 3: I don't think so.

Team FACE Member: How important is it to you to recognize emotional cues in your daily interactions?

Subject 3: I lost my vision 14 years ago. I miss seeing people's faces. To see if they are angry, happy, or disgusted. I don't care about the color of the person or if they are attractive or not. It used to make a big difference to me. I want to see if what I said upset them or made them happy or sad. Sometimes I can tell with tones but not all the time. But anybody can fool anybody.

Team FACE Member: Can you rate on the scale of 1-5 with 5 being the highest, how important it is to recognize facial cues?

Subject 3: 3.5

Team FACE Member: Can you rate on the scale of 1-5 with 5 being the highest, how interested are you in a portable device that can recognize faces and emotions?

Subject 3: 5

Team FACE Member: What is the best way for the device to convey expression cues to you? Auditory, beeps and chimes, or vibrations?

Subject 3: Audibly. Unless I knew that certain chimes or beeps meant certain things like a change in volume. It would be useful but I would have to learn how to do that.

Team FACE Member: Would you like the option of switching between auditory and chimes?

Subject 3: Yes.

Team FACE Member: Which methods would pose significant inconveniences for you? For example, if you chose audio feedback is there a case where it would distract you or interfere with your daily life?

Subject 3: Well depends on where it is coming from. Is it coming from a portable device on me all the time or on the computer or where it is?

Team FACE Member: In this case it is a removable device. Something that is on you.

Subject 3: When having a conversation with somebody and it keeps telling me expressions, there should be a way to turn it off. If I am in a quiet place, I could have earphones so only I can hear it. Places where they tell you to turn off your cell phone.

Team FACE Member: That's why we thought of the vibration feedback. In this case you would be wearing a belt or other device that would vibrate

depending on where and strength.

Subject 3: I do not like the idea of the vibrating. I don't think that I would enjoy experiencing that on a daily basis. I think it would be annoying but I don't use any devices that vibrate.

Team FACE Member: How comfortable would you be using a cane mounted with a camera attached to a laptop in a backpack in public?

Subject 3: I don't mind being seen with it. I think the whole way of doing it sounds very cumbersome. If I am standing there and using it for a short amount of time it would be fine. But if I am doing that for life it would be annoying.

Team FACE Member: Could you rate how comfortable you would be with this device on the same scale?

Subject 3: 1 or -1. I would not like it.

Team FACE Member: How would you like it?

Subject 3: Like you said, something portable that can fit in my pockets like a cell phone or around my waist. Something less obtrusive and fewer things to carry and maintain. You are talking about using this thing on a regular basis like all the time.

Team FACE Member: Yes. What if it is something that is removable and carry with you not mounted on a cane?

Subject 3: I would not mind that.

Team FACE Member: If you had the original device, would you wear it while interacting with visually impaired people?

Subject 3: I would not.

Team FACE Member: If you had the original device, would you wear it while interacting with sighted people?

Subject 3: I might...if I had a strong need to know their emotions. No, mostly because I don't use anything like that right now. I feel like it would make people shy away from me because it is big and cumbersome and they think that they are being recorded.

11.1.4 Subject 4

Team FACE Member: So I have a few quick questions that are probably the easier ones that we can begin with. You don't have to answer any of these if you don't want to. They're just a demographic questionnaire. Can I ask your age?

Subject 4: 45.

Team FACE Member: Ok. And your gender?

Subject 4: Male.

Team FACE Member: And level of household income. You can also not answer this if you're ok with that. We have from A through F. There are choices A) $0 - 19,999; B) 20,000 - 39,999; C) 40,000- 59,999; D) 60,000 - 79,999; E) 80,000 - 99,999; and F) over 100,000 a year

Subject 4: D

Team FACE Member: How about ethnicity? We also have A through E in terms of choices: A) American Indian/ Alaskan Native; B) Black or African American; C) Asian or Pacific Islander; D) Hispanic/ Latino; E) White; F) Other or Not reported

Subject 4: B.

Team FACE Member: And level of education?

Subject 4: B.S. Degree.

Team FACE Member: Great, let's move to the first question. First question I

have here is how do you describe your social interactions with other blind people?

Subject 4: I would describe it as a very good interaction, because we can identify with someone with same disability.

Team FACE Member: Do you ever think there's anything missing?

Subject 4: Not really, the only thing that is missing...you guys put in your head people's expressions. People really rely on that a lot. If someone is really happy or sad, you can hear it in your voice. But facial expressions are good to help you with those kind of things as well. It's good to know before you begin to talk to them, what they have in their face, but everything else, they are great

Team FACE Member: So how do you describe your interactions with sighted people?

Subject 4: It's very good, no difficulties. Sometimes, it's a little frustrating. They may not understand exactly what our particular needs are because of our sight loss, but not with everyone.

Team FACE Member: Do you mind me asking if you were blind at birth or if you became blind later?

Subject 4: I was diagnosed with RT at age at 7 and lost majority of vision in high school. But at least I had the opportunity to be able to put in my mind's eye an image, when someone is describing something to you

Team FACE Member: Do you currently use any assistive technologies or devices that help you in your social interactions?

Subject 4: Computer is very important, JAWS, Assitive tech. works out very nice, just can't see myself without it, for this kind of thing anyway.

Team FACE Member: Do you carry anytihng around with you?

Subject 4: Digital recorder. You would be able to grab pen and paper to

write things down. I would bring digital recorder for when people are trying to share information with me

Team FACE Member: Using a scale of 1-5, 5 being the highest, rate how important it is for you to recognize emotional cues in your daily social interactions.

Subject 4: Since I don't have anything to help me to do that right now, I would have to say it's not all that important, so maybe 3.

Team FACE Member: Rate your interest, on a scale of 1-5, 5 being the highest, in a portable device that could recognize faces and convey identities to you.

Subject 4: I guess if I had something to use for facial expressions...How often would I use it? Probably a 4.

Team FACE Member: Rate your interest, on a scale of 1-5, 5 being the highest, in a portable device that could recognize and analyze facial expressions and convey these cues to you.

Subject 4: 4.

Team FACE Member: What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.?

Subject 4: Vibrate, because if you are with some and all of a sudden you are chiming and beeping and all sorts of stuff, so if you had a vibrating device and it would vibrate, people would be like whats that.

Team FACE Member: What if through earplugs?

Subject 4: Then you have think about maybe a Bluetooth device, like off a cellphone. that would be great. People would usually have a earpiece device. I have an earpiece to speak person smiling. Person frowning to you right now, so if I could hear it myself without having myself hear it, that would be good.

Team FACE Member: Would you rather have sound or vibrate?

Subject 4: As long as we know what beeps mean, that would be nice. If we could press a button and hear what person's expression is, directly, that would be even more helpful. If you can't, then chimes are fine.

Team FACE Member: Do you usually bring a cane around?

Subject 4: Yes, it's like our vision.

Team FACE Member: On a scale of 1-5, 5 being the highest, how comfortable would you be carrying a cane mounted with a camera connecting to a laptop in a backpack in public?

Subject 4: The smaller the item, the better off you are. They got these little cameras which is very inobstrusive, not a problem. To convey info, you need to connect thorugh laptop. It's quite a bit to get information. If you ask me if carrying it around would be worth having that information, then I would say no.

Team FACE Member: Do you think it's useful?

Subject 4: If its not obstrusive, 3.

Team FACE Member: So you think expressions are more important?

Subject 4: Of course. You usually know who the person is - we have voice signature. You know who the person is. Ex. I know that's Team FACE Member. But is (Team FACE Member) smiling, is she unhappy today? That's more important. People who we interact with on a day to day basis are people who we already know. If we are interacting with a stranger, then we don't have voice signature, so we would want to know as much as we can about that person's expressions.

Team FACE Member: What would you give this out of a 1-5?

Subject 4: 1.

Team FACE Member: What if you took the laptop out of it, and it all ran off of your existing cell phone.

Subject 4: Oh yes! That would be terrific!

Team FACE Member: And we have the last question: If you had such a device, would you wear it when interacting with other visually impaired people?

Subject 4: I think I could use if for everyone, not only blind, but blind as well as sighted.

11.1.5 Subject 5

Team FACE Member: What's your age?

Subject 5: I'll be 59 on Tuesday.

Team FACE Member: What's your gender?

Subject 5: Male.

Team FACE Member: Level of income?

Subject 5: I'm retired now, but before retirement $100,000+.

Team FACE Member: Ethnicity?

Subject 5: Italian-American.

Team FACE Member: Level of education?

Subject 5: Post-Masters.

Team FACE Member: How do you describe your social interactions with other visually impaired people? Sighted people?

Subject 5: They're pretty normal, pretty even, has nothing different at all.

Team FACE Member: Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use? Rate how effective this device is from 1-5, 5 being the highest.

Subject 5: My cell phone speaks the numbers out loud so I know who's calling me, I have a talking watch that tells me what time it is, I have a talking calculator, and I'm really not computer literate yet, I'm scared to say [] I'm supposed to do that soon; I'm not a Braille reader, but I'm an auditory listener/learner type person, I'm also a speaker and a class teacher, so socially interacting is not an issue at all. I also use a cane for travel obviously. I think they're all a 5 otherwise I wouldn't use them. I like to maximize my stimulation in terms of sensory to replace vision with those things, so they help me a lot.

Team FACE Member: Using a scale of 1-5, 5 being the highest, rate how important it is for you to recognize emotional cues in your daily social interactions.

Subject 5: I think it should be a 5, but since I can't see visual cues or body language, I probably don't actually realize it's a 5 although I believe it should be a 5. But I'm pretty tuned in to people's voice and so forth, and I'm very, very conversant in intercommunicating so I can usually pick up if someone's not feeling maybe too well or vice-versa.

Team FACE Member: I watched one of your videos on YouTube and you're very inspirational. So you lost your vision, you weren't born blind, right?

Subject 5: Correct.

Team FACE Member: Okay. So do you think that affects how you deem how important emotional cues are in your daily life because you grew up being able to recognize?

Subject 5: Yeah, that's a great question. You know you don't really enhance the number of sense, you use it more. Like people think I have super-hearing, but my hearing really isn't that great. I have tremendous visualization in my mind; every day I walk around 24/7. I have a mental camera in my eyes, a picture of

everything that used to be. So I remember what people look like, and landscapes, and colors and so forth. So I just turn on a virtual movie in my mind's eye 24/7, so I never see darkness.

Team FACE Member: How long has it been since you lost your sight?

Subject 5: About 26 years. My eyesight for 10 or 12 years slowly dissipated from driving and playing sports in my early 20's, around your age, down to my early 30's where I couldn't see. So it took a long time to slowly go away, it wasn't an overnight thing. That's typical of retinitis pigmentosa, by the way.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize faces and convey identities to you.

Subject 5: That's a tough question. I deal so well with people, but if I can recognize emotions...I'm an input kind of person the more input I get, the better it is. Especially if I were speaking on stage for a presentation, it'd be nice to see the smiles on people's faces.

Team FACE Member: Right, but this question is specific only to facial recognition.

Subject 5: Okay. It's not a 5, believe it or not, I wouldn't say it's a 1 either. I'd say a 3.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you.

Subject 5: Probably the same, 3-4.

Team FACE Member: Do you rank the expressions higher than facial recognition?

Subject 5: Yes.

Team FACE Member: We were very surprised because some of the interviewees we had preferred facial recognition a lot more than facial expressions.

Subject 5: Out of curiosity, were those people blind since birth?

Team FACE Member: Yes, I believe so.

Subject 5: That's what I would've thought. People like me who've seenfor facial recognition, that's a curiosity of totally blind people, to know what your faces looks like. For me, it's more the expression of their feeling because I know what their face looks like.

Team FACE Member: And does that go back again to because you weren't born blind?

Subject 5: Probably, I believe that would be the case.

Team FACE Member: What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.? So for example, if someone's smiling, would you like the device to

Subject 5: I'd rather it speak it to me than some beep or something, I feel like I'm a robot.

Team FACE Member: Okay, so you'd rather have audio?

Subject 5: Yeah, I'd rather it be spoken, like some verbal cue, almost like a descriptive video.

Team FACE Member: Okay, do you think that might get annoying?

Subject 5: It's almost like, "'Hey, there's a nice-looking girl, she's got beautiful eyes, she's smiling, she's holding out her hand to you."' That's much better than "'beep, beep, beep."'

Team FACE Member: But would the audio distract you from conversation? Because it'd actually be spoken to you instead of just something in the background.

Subject 5: No, I think it'd be kind of cool. It's almost like someone giving me a head's up.

Team FACE Member: Okay.

Subject 5: And by the way, just as a little aside, you may want to look into descriptive voice video for the blind. It's really interesting. You know those videos that have subscripts for deaf people? There are actually videos for blind people, movies that are out there that you can get. It'd be kind of cool for you. And you watch it, and it's like you know when you're in a movie and someone's whispering to you and people get mad? But it actually tells you what's going on. Like in *Dumbo*, Dumbo the elephant now is flying in the air and singing and landing on the ground. As you hear the main thing, you hear the little voice behind it describing everything, really cool.

Team FACE Member: So did you say that was descriptive voice video?

Subject 5: The NFB has those. In fact, it's really cool for people who can see, it's really neat.

Team FACE Member: So for the feedback systems that I asked you about, can you name some methods that would pose significant inconveniences to you? Like are there any feedback methods that you absolutely wouldn't want on the device?

Subject 5: Personally, I'd rather it be a voice. Beeps and so forth make me feel dehumanized, robotized, and almost like some kind of computer chip. I'm a communicator, I want to be just like any other person, so voice cues and all to

me would be a little like Pavlovian, a little like Pavlov ringing the bell and I'll salivate. I'd feel just a little puppeteered like that.

Team FACE Member: And would you want the option to be able to turn off the voice, say you're in a conversation and you think the voice is distracting? Do you want the system to be ongoing, or do you want to have control over?

Subject 5: I think I should have control over it because, yeah, it'd be kind of corny. Are most people wanting voices or beeps and sounds?

Team FACE Member: Most people want the beeps and sounds because they thought the voice would be too distracting.

Subject 5: You know what's interesting I'll bet those people who said it were blind since birth.

Team FACE Member: Yeah, you're right. Subject 5: I hate to say it, but that group is much more programmed. They're more following prompts and they're really more follow-the-leader, follower-type people to cues because they have no visual cue. And describing what something looks like to a person who's blind is sort of irrelevant because they have no reference point. You're going to see the people with RP (retinitis pigmentosa), my postulation would be, the people who used to see, are going to answer like me. The people who've never seen would want to be beeped at in this way, that's just my hypothesis.

Team FACE Member: That's really interesting. On a scale of 1-5 (5 being the highest), how comfortable would you be carrying a cane mounted with a camera which connects to a laptop in a backpack in public? So basically what we're trying to do is mount a camera on the white cane, and this camera gathers all the input from the surroundings so facial recognition, facial expression, and then communicates it to you through the device that you wear.

Subject 5: So this would be my eye and tell me what it's seeing.

Team FACE Member: Exactly.

Subject 5: That's pretty cool. I'm pretty comfortable with myself so I'd be very comfortable. I'd be a 5 on that. Now, I'd also have to learn how to work a computer finally.

Team FACE Member: If you had such a device, would you wear it when interacting with other visually impaired people? With sighted people?

Subject 5: Yeah absolutely, why not? If I decided I needed to use it, I'd use it period.

11.1.6 Subject 6

Team FACE Member: How old are you?

Subject 6: 40.

Team FACE Member: Gender?

Subject 6: Male.

Team FACE Member: Level of income?

Subject 6: $100,000.

Team FACE Member: Ethnicity?

Subject 6: White/Caucasian.

Team FACE Member: Level of education?

Subject 6: Master's.

Team FACE Member: How do you describe your social interactions with other visually impaired people? Sighted people?

Subject 6: I would say they're positive. Are you looking for more?

Team FACE Member: No, not necessarily, but do you feel comfortable interacting with sighted people?

Subject 6: Absolutely.

Team FACE Member: Just as background, were you born blind or did you lose your sight later on in life?

Subject 6: I lost my sight when I was 19 years old.

Team FACE Member: Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use?

Subject 6: Yeah, I use a screen reader, specifically JWS, I use a Braille display. For social media, I use Twitter.

Team FACE Member: Do you use the white cane to get around?

Subject 6: Yes, I do.

Team FACE Member: Can you rate how effective theses devices are from 1-5 (5 being the highest)?

Subject 6: I would say 4.

Team FACE Member: Okay, for all of them?

Subject 6: Yeah.

Team FACE Member: Always room for improvement?

Subject 6: Exactly.

Team FACE Member: Using a scale of 1-5 (5 being the highest), rate how important it is for you to recognize emotional cues in your daily social interactions.

Subject 6: I would say 4.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize faces and convey identities to you.

Subject 6: A 3.

Team FACE Member: Can you explain why?

Subject 6: Because over the last 20 years, I've been able to master alternative techniques related to the way someone is breathing to help me gather that information.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you.

Subject 6: I would say a 2 and the only reason I would say that is because I've never imagined of such a device. So to see how useful that would be may be interesting, but you're asking me if I'm interested in something I've never thought about is like asking someone in the 70's if he'd be interested in a cell phone. It's like a "'What?"' So I'm not really sure how that would work. I mean, I'd be interested in seeing how it would work and how it could function, but is my desire that strong to have one? Not until I saw it.

Team FACE Member: Okay. So if we were able to build this device, would being able to recognize faces or facial expressions be more important to you? Or are both pretty much equal?

Subject 6: I would say probably the expressions would be more interesting. Because if I knew who they were by their voice, I'd feel like their isn't any news that I wasn't getting.

Team FACE Member: Okay, so you just ranked expressions lower because you hadn't thought about such a device existing before, right?

Subject 6: Correct.

Team FACE Member: What would be the best way for a device to convey

facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.?

Subject 6: I would say that the vibrations would be useful because the other methods may interfere with the conversation.

Team FACE Member: So you think that audio, like being spoken to, beeps, and chimes would interfere with the conversation? And would be distracting?

Subject 6: Yeah, I think that that's what I would imagine.

Team FACE Member: So those would be inconvenient.

Subject 6: Right. You left out Braille. Braille would be another way to get that information and wouldn't interfere with the conversation.

Team FACE Member: Okay, thanks, that's a great suggestion. On a scale of 1-5 (5 being the highest), how comfortable would you be carrying a cane mounted with a camera which connects to a laptop in a backpack in public? So basically what we're trying to do is build a device where we mount a camera on a white cane and the camera collects information about the surroundings, like people's faces, people's facial expressions, and then sends the information to the blind individual through, say, Braille feedback or vibration feedback.

Subject 6: I think you'd have more luck with using the cell camera technology.

Team FACE Member: Instead of a camera mounted on a cane?

Subject 6: Yeah. The iPhone and Nokia phones all have the cameras on the front and the back and I think that that technology, if you can connect to a laptop, be able to use...

Team FACE Member: So that would be a lot more convenient because it's smaller?

Subject 6: Yeah.

Team FACE Member: Okay, the only concern we have with that was that, when we mount a camera on a cane, the cane is pretty much relatively eye-level to adults. But with an iPhone, for example, the blind user would have to position the camera in order to receive the proper feedback.

Subject 6: Yeah, but you're also assuming that I'd be standing and having this conversation, where I think what I would like with this technology would be more at a meeting where I was sitting down. And when I'm sitting down, I'm not using my cane.

Team FACE Member: Okay. So the camera on a phone would be a lot more flexible in those meeting situations.

Subject 6: Correct.

Team FACE Member: If you had such a device, would you wear it when interacting with other visually impaired people? With sighted people? (For the cane device instead of phone device)

Subject 6: I guess so.

Team FACE Member: And same with the phone camera?

Subject 6: Yes.

Team FACE Member: But you would prefer the phone camera?

Subject 6: Yeah.

Subject 6: Ray Kurzweil has invented optical character recognition software, so I can take a picture with my cell phone and it will read to me what the document is. Optical character recognition is called OCR. He's also working on OSR which is optical scene recognition where it would be able to tell you you're in a park, you're in a restaurant, you're in a building, you're in a cafeteria at a certain table. With the OSR software, he'd be able to capture the similar

information that you wish to capture.

Team FACE Member: Okay, thank you for that contact.

Subject 6: The one thing you didn't ask is: are people working? And that's a critical issue with people with disabilities. And it's relevant because if you can't afford this technology, then who's going to buy it supply and demand. So are you working is capturing that type of information, I think that might be interesting.

11.1.7 Subject 7

Team FACE Member: What is your age? Subject 7: 64

Team FACE Member: Gender? Subject 7: Female

Team FACE Member: Level of income? Subject 7: Above $80,000

Team FACE Member: Your ethnicity? Subject 7: White

Team FACE Member: Level of education? Subject 7: I have a Master's Degree.

Team FACE Member: How do you describe your social interactions with other visually impaired people? Subject 7: Easy, I don't have any problems communicating with people with visual impairments.

Team FACE Member: Okay, and with sighted people? Subject 7: Sometimes a little iffy, most of the times they're fine. You know, I'm 64 so I've had time to learn how to do it. Most of the times I'm comfortable, sometimes I'm not.

Team FACE Member: Were you born blind or did you lose your sight later in life? Subject 7: I was a preemie, so I was essentially born blind. I was blind immediately after birth.

Team FACE Member: Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use? Subject 7:

I use a cell phone, I use a Braille note to take notes, keep track of things, and addresses and information. I'm a big email person, I use a computer. I guess that's all.

Team FACE Member: Could you rate how effective this device is from 1-5 (5 being the highest). Explain your rating. Subject 7: Effective for what? For assisted living interaction? Team FACE Member: Exactly Subject 7: A cell phone would be a 4, the Braille note would be a 3, and the computer would be a 4.

Team FACE Member: Using a scale of 1-5 (5 being the highest), rate how important it is for you to recognize emotional cues in your daily social interactions. Subject 7: A 4 because you don't know if someone is upset by what you're saying or reacting negatively. You could just make the situation worse by not recognizing that and changing your communications accordingly.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize faces and convey identities to you. Explain. Subject 7: I don't know if I'd like to take something like that with me all the time, so I think I'd give it a 2.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you. Explain. Subject 7: I would be interested to see how it works and I think it'd be fabulous for kids. You could correlate what they recognize with tone of voice and body language and stuff that they...I think it'd be a great teaching device, so I think I'd give it a 4.

Team FACE Member: You said for kids? Subject 7: Yeah, I think it'd be a good thing. I would be interested in it too, but I think it'd be a great way to

teach blind people social skills because those are the harder things for kids to learn.

Team FACE Member: And so you would be more interested in a device that would convey facial expression as opposed to facial recognition information? Subject 7: Yes.

Team FACE Member: What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.? What methods would pose significant inconveniences to you? Explain. Subject 7: I think it should as unobtrusive as possible, so maybe you could do it tactically so that people wouldn't know that it was doing that. Like if you had something in your hand or a watch strapped on your wrist, and it vibrates at different ways and different spots, I think that plays the most effective way to do it and it wouldn't make people feel uncomfortable because they wouldn't know you were doing it. Team FACE Member: Okay, but for the audios and beeps and chimes, we were envisioning an ear piece kind of device, so those around you wouldn't be able to hear the conversation. It'd only be the person using the device. We just weren't sure whether this would interrupt the conversation. Subject 7: It would. That would be distracting.

Team FACE Member: On a scale of 1-5 (5 being the highest), how comfortable would you be carrying a cane mounted with a camera which connects to a laptop in a backpack in public? Explain. Subject 7: A 1, sorry. I'd give it a 1. I'm not a cane user, I'm a dog user. You could put it on a dog's harness. Team FACE Member: Would you use it if it were on a dog? Subject 7: I doubt it. I would be very interested in trying it out.

Team FACE Member: What kinds of devices would you be interested in

seeing? Subject 7: I would be interested using something as unobtrusive as possible. Team FACE Member: So maybe like a cell phone camera?

Subject 7: Yeah, that'd probably be a really good idea. Make it an app for the iPhone or Droid, I think it'd be cool.

Team FACE Member: If you had such a device, like the iPhone or Droid, would you use it with blind people or sighted people. Subject 7: With blind people, no, I probably wouldn't. Blind people are used to personal information (tone of voice).

Team FACE Member: Can I just ask you if you're currently working right now? Subject 7: I was kind of working. I'm unemployed but I have some contract work, so it's not steady but occasionally I'll be lucky enough to get some writing assignments.

11.1.8 Subject 8

Team FACE Member: How old are you?

Subject 8: I am 66.

Team FACE Member: Gender?

Subject 8: Female

Team FACE Member: Your level of income?

Subject 8: I'm retired. [Before that] it was below $40,000.

Team FACE Member: Your ethnicity?

Subject 8: I'm White.

Team FACE Member: Your level of education?

Subject 8: I have a Master's degree.

Team FACE Member: How do you describe your social interactions with other

visually impaired people?

Subject 8: No problem.

Team FACE Member: And with sighted people?

Subject 8: No problem.

Team FACE Member: Do you currently use any assistive technologies (devices) that help you in social interactions? If yes, what do you use? Rate how effective this device is from 1-5 (5 being the highest).

Subject 8: No.

Team FACE Member: Do you use the white cane?

Subject 8: Yes.

Team FACE Member: And on a scale of 1-5, how effectively would you rank this device, with 5 being the highest?

Subject 8: 5

Team FACE Member: Using a scale of 1-5 (5 being the highest), rate how important it is for you

to recognize emotional cues in your daily social interactions.

Subject 8: A 5. I think it's important to understand where people are coming from, so you can have a better connection with peopleif you can sense, if you know they're sad, or happy, or uncertain, or troubled, or whatever.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize faces and convey identities to you. Explain.

Subject 8: Because of the kind of vision I have, I have just light perception at this point, I don't think that that kind of device would work for me. I mean, I don't know what the technology would be and so forth, so are we pretending

that this thing could help me?

Team FACE Member: Yeah

Subject 8: So it's as if the potential is there, and you're asking me if I would use it?

Team FACE Member: Exactly, if it had a facial recognition feature

Subject 8: Yeah. I think I would, yeah.

Team FACE Member: Could you rate your interest, then, 1-5?

Subject 8: I think I'd do 5.

Team FACE Member: Okay, and could you explain why again?

Subject 8: I pick up all kinds of cues right now by voice, by how people sound. So if that person had facial, that'd be all that much better.

Team FACE Member: Rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you.

Subject 8: So recognize and analyze? You mean, it would say to me this person is smiling?

Team FACE Member: Right, exactly. Or it would have a beep or audio clip that correlates to a certain emotion.

Subject 8: So I wouldn't see the face, but the device would clue me in.

Team FACE Member: Exactly. Even for facial recognition, if a person you knew was approaching, then the device would ideally say, Such and such person is approaching. And then for the facial expression feature, when you're engaged in a conversation, for example, if someone smiles you may hear the device say, So and so is smiling, or you may hear an audio clip that is correlated to someone smiling.

Subject 8: Here's my comment. I think that, depending on what the device looks like, I think that a device that was very, very obvious would change the interaction with somebody because it would be the focus of the interaction, rather than just the two people. The other person would see that, and that would change the interaction I think. And the fact that I'm visually impaired and, now, basically blind of course, that's just who I am. And so if I interact with people, that's who I am. So I don't think that gets in the way for me because it's just me. But if there were this device that was sort of like a big TV or something else wearing on my shoulders, then that I think would change what going on. I don't know.

Team FACE Member: No, I understand that. Right now, we're trying to mount a camera on a white cane because you always have your white cane with you or, in most cases, the camera goes with you. And we don't think it'd be that obtrusive. Originally, we had one that mounted cameras on glasses for example, but not all blind people wear glasses, and we were concerned with the weight of the camera on the glasses. And then there was a suggestion to make a chest mount and put a camera on that, or have a hat, but then that's very obvious, it's not very discreet. So right now, we're focusing on mounting a camera on a white cane because we think it'd be most convenient and the most discreet.

Subject 8: And have you done any kind of mobility training? I know you've worked with some visually impaired. Have you used a white cane? Do you know how it's used?

Team FACE Member: We've seen how it's used but we, ourselves, have not tried using it.

Subject 8: Okay, because when I use a cane, I am moving it constantly. I'm

moving it back and forth, sometimes I'm holding it one way, sometimes another way. Sometimes it's slantedit just totally depends.

Team FACE Member: Right, so there's a lot of variability.

Subject 8: So I just have questions about the type of information it'd be very interesting to see how that could work.

Team FACE Member: Right, well that's one of the problems that we know the white case presents. One of the solutions we're trying to implement is putting some sort of marker that you can feel on the back of the cane to know when the camera is facing forward.

Subject 8: [Approval]

Team FACE Member: So otherwise the camera could be turned in any direction. So that's one of the things.

Subject 8: Exactly. And it wouldn't be level unless it were on sort of a swivel and had a weight that would always stay level to the ground, just like a compass always points north. It would have to stay. That would be very interesting. Plus it's down lower, and so that's an interesting problem.

Team FACE Member: The other feature of the white cane was that, if we mount the camera on top of the white cane, it's pretty much eye level with other people or around their face.

Subject 8: The straight cane that someone uses the top of it should come about 3 inches down from your neck. So between the shoulders, in front of the sternum, if you want to get the right size, it's really there. So it really depends on the height. If you're talking to a guy, then a girl might be up level at that height. But a short person would be up at that tall guy's belly button.

Team FACE Member: So we have to think about that too. When we went to

NFB, we saw people using the white cane, and that's why we thought that it'd be pretty much eye-level. But I guess that could vary a lot.

Subject 8: The people at NFB they use these very long canes, and not everybody uses those. Those are really up high and way out there in front, but there are a lot of people that don't use that kind. I think more people use the kind I use, and then there are visually impaired people that have a white cane that's like a support cane. It's like a cane that anybody would use, like an older person would use for support, but it's white because they're visually impaired. But they also need it for support, so it's white. So that'd be down at the waist. So there are a lot of different sizes. The size I use is a 50 inch. And I'm holding it in front of me at my belly level as I use it.

Team FACE Member: That's great information, we weren't aware of that.

Subject 8: Yeah, so if it's standing up, it's only 50 inches. And I'm 5'6"', so I'm not short.

Team FACE Member: So could you rate your interest, on a scale of 1-5 (5 being the highest), in a portable device that could recognize and analyze facial expressions and convey these cues to you.

Subject 8: So we have one that could recognize faces and one that could recognize emotional facial features?

Team FACE Member: Right, ideally we'd like to combine both aspects, but we wanted to get a sense of what was more popular.

Subject 8: So the face would be, "'Okay, this is my friend."' And the other would be, "'Oh, my friend's crying."'

Team FACE Member: Exactly.

Subject 8: Yeah, I think it'd be interesting to know those things but, as soon

as you start talking to somebody, you know those things.

Team FACE Member: Even with people you don't know?

Subject 8: Well, yeah, that's true.

Team FACE Member: Because the feature with recognizing faces would only work for those you know because we would have to input their face photos into the software, whereas for the facial expression, that analysis should be applicable to anybody, not just for people you know.

Subject 8: So you'd take pictures of people smiling and call it smiling?

Team FACE Member: Something like that. The algorithm that we're working with, from what I understand, analyzes how much the facial expressions change. It focuses on the eyebrows, on the corners of the mouth, that type of stuff. It's live, it keeps on going, it's in real-time.

Subject 8: I think that'd be interesting. I'm not saying I'd use one, but I think it would be fascinating to see that kind of thing.

Team FACE Member: So on a scale of 1-5, how would you rate it, with 5 being highest?

Subject 8: How would I rate it as to whether I would use it or if it's desirable?

Team FACE Member: Yes.

Subject 8: I think it'd be a 5.

Team FACE Member: Would you be more interested in the facial recognition or facial expression?

Subject 8: I think that the facial expression thing would probably be overall more helpful because the other one, you'd have to take pictures of all your friends and get them all in there.

Team FACE Member: And could you explain why you rated the facial ex-

pression analysis feature a 5?

Subject 8: Because it would be useful in any situation, you could use it whether with the people you know or without, whereas the other one would just tell you that that's John walking up. This one, if you know this guy, if he's going to say hello and you know who he is...you could figure that out because you know it. He's going to say hello and you know who he is. And if it's one that gives you the emotion, you can then see how this friend is doing. And, of course, you could do that with anybody.

Team FACE Member: Just to clarify, are you completely blind?

Subject 8: I have light perception. When you go for an eye exam and you have 20/20, then that's perfect. If you get to 20/200, that means you're legally blind, but it depends. If people get up close enough, you can see whether they're smiling or recognizable. And from then you go to count fingers, you know the doctor will hold up and see how many fingers do I have, and how many feet the doctor's holding this up in front of you. So you go from that to counting fingers and then the next one would be hand motions. The doctor would say, "'Can you see my hand moving?"' And that's hand motions. And then the next one is light perception where you can tell if the lights are on or off. And, of course, after that it's no light perception. Those are the variations. I never had perfect vision, but I had enough vision where they didn't know, when I was a kid, that I had a problem until I couldn't see the blackboard in school. So I must have had really good vision back then, not perfect. And it has gradually diminished, so now I'm to where it's light perception. And sometimes I can't tell. If I come in from a sunny day, I can't tell if the light is on or off inside. And unless I get up close to a lamp, I can't always tell if the light is on.

Team FACE Member: So can you see general shapes?

Subject 8: No, I can't see that anymore. So I wouldn't be a candidate for any of these things I'm sure. Because I think if you have anybody who's going to test these things, they've got to have enough vision to be able to. Unless it's all auditory, so it really depends on what you end up working on.

Team FACE Member: Yeah, well that was my next question. I was wondering what would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.?

Subject 8: I don't know, but I think it would need to be something that wasn't distracting, that could give you the information and not take away from the interaction. And not be observed by the other person.

Team FACE Member: So ideally, the user would be wearing an earpiece, so all the audio (if audio was chosen) would be streamed in through the earpiece. For facial recognition, the only option is audio, but for facial expression, there are a number of options. There could be, for example, vibration. A previous team made a haptic feedback belt, which is a belt that has vibrations in different areas. So, for example, we could program if someone were happy, then the belt would vibrate on the left side.

Subject 8: Now I think that would be good, I think that's what I would prefer.

Team FACE Member: As opposed to audio, or beeps and chimes?

Subject 8: Yes. I wouldn't want anything to disturb my audio, that's where I get most of my input. I wouldn't want anything covering up what I'm already getting, so I think the vibration thing or buzzer around the waist would be great.

Team FACE Member: Okay, so for vibration, would you be more interested in a belt around your waist or maybe a wristband, or do you have any ideas?

Because we want it to be comfortable.

Subject 8: Well for the wristband, you could have front and back sides, or maybe it could be different kinds of vibrations, such as steady beats or pulsing. It'd have to be something that was not too obvious. I think the belt would be good, except it may get hot in summer. And it'd be bulky, it could make your clothes bulky. But this wrist thing might work.

Team FACE Member: On a scale of 1-5 (5 being the highest), how comfortable would you be carrying a cane mounted with a camera which connects to a laptop in a backpack in public?

Subject 8: Well, I wear a backpack anyways when I go to yoga class, I put my yoga mat in it. Or if I'm heading somewhere, wearing a backpack is no big deal. It would depend on how heavy it is, how hot it is, that sort of thing. But I'm telling you, that whole cane-camera thing makes no sense to me because of the way you have to move the cane around. In order not to fall in a ditch and to find what you're doing, the cane is constantly moving. And you're going to take great pictures of the treetops. It's not going to be steady, it's totally moving. And it's slanted. You may see people standing around with their canes up around their head if they're standing and talking to you, maybe that's what you're thinking of. But if you're walking, it's different. So if you're talking about stationary stuff, it's a possibility.

Team FACE Member: Right now we're just focusing on stationary.

Subject 8: Oh okay.

Team FACE Member: Yeah, moving would present a lot of obstacles. And ideally, we'd like to get around to moving, but what we had pictured was when you're holding your white cane and someone enters the room, then the camera

would be able to pick up who it was and recognize them. And when you're involved in a conversation, ideally, that'd be stationary too. So the only problem that we're dealing with is whether the camera is facing forward now and the height level. Right now, we're not so much focused on whether the camera can pick up things while the camera is moving around because that's the next step.

Subject 8: That's interesting. I should tell you that I worked for almost 20 years as a rehab teacher with Department for the Blind in Virginia and I didn't teach mobility, you had to be sighted to teach that but I took mobility training with a cane, and nobody in my agency ever used those tall canes that are as tall as your head. And this was a state agency. I mean, I know they exist, but if you're...it depends on the kind of people you're working with. If you're talking aboutjust like there are all kinds of people, there are all kinds of blind people, and some are out there. They'll ride the subways and they're out there by themselves and very independent. And the National Federation of the Blind is made up of those kinds of people who are professionals and out there at their cocktail parties. And they probably might use that kind of thing. So there are lots of people that use these canes that I'm talking about that are no higher than your shoulders, kind of to right below where your sternum is and your breastbone starts, about 3 inches down from there is the ideal height. And then you have the ones that are down low, like support canes.

Team FACE Member: So would you have any other ideas of how to mount a camera?

Subject 8: The only thing I can think of would be a lovely pendant around the neck. But still that's too low. It might work for a tall guy, but is he going to wear a necklace? What about a short woman or a shorter person? I don't know.

I know they have these things like goggles you wear around your head.

Team FACE Member: But would people actually wear that out into the public?

Subject 8: No, that's just it. In a way, that puts you in a separate category. And when you're trying to socially connect with people, what you don't want is to make yourself different. You're already different enough because you've got your cane and you can't see them. So you don't want to add to that so, of course, that's why I don't think that really worked years ago.

Team FACE Member: There was a suggestion of using a camera on a cell phone and, I know that's a good idea but, again, there's the problem of finding just the right height of where to place the camera to get maximum input.

Subject 8: Yeah

Team FACE Member: But that would be one of the smaller options and the more discreet option.

Subject 8: Yeah, but they'd have to hold that camera up. And, again, that's a little odd. I know, maybe people have their cameras in their hands all the time (on their phones, I mean) because I know people text all the time. But what if they have it pointed towards the person they're talking to, it might make the other person a little nervous. Like, "'Why are you taking a picture of me?"'

Team FACE Member: So I know you express a lot of concern about the white cane because of the height, so could you just rank, on a scale of 1-5 (5 being the highest again) whether you would use this sort of device in public?

Subject 8: Yeah, I think it'd be interesting to try it. Because I'm not a real tech person, I have doubts about that sort of thing, but I think it'd be interesting to try.

Team FACE Member: So what would be your rating?

Subject 8: If I wanted one of these?

Team FACE Member: Yeah

Subject 8: I think I'd give it a 5 as to the possibilities could be really good and they could be terrible. But why not think that they'd be good? So I'd say a 5.

Team FACE Member: Would you be comfortable switching over to a taller cane or not really because you're already used to your cane?

Subject 8: Oh no, I wouldn't use a taller cane.

Team FACE Member: And do you think most people are like that if they are already used to their cane?

Subject 8: Most people who have trained with the cane that comes up just to the sternumI don't know whether people would switch. I've never seen anybody use those long canes. I think you should talk to some mobility instructors, not just the NFB people. Because that's one kind of person, a professional. And their view is different. And because I worked in the field for 20 years and I worked with a lot of blind people who have got mobility training, and I never saw anyone use those tall canes, and I worked with the population from 18-100that's why I think you should talk to somebody who's training people to use the cane in the field of just the average person. That person is called the orientation and mobility instructor. You could call wherever the Maryland agency is closest to you. Just explain your situation to them and see if you could talk to them, and it's called O and M, ask to speak to an O and M person about some questions that you have.

Team FACE Member: If you had such a device, would you wear it when

interacting with other visually impaired people?

Subject 8: I would say that 99% of my friends are sighted and, at this point in my life (now that I'm retired), the blind people that I'm friends with are at a distance because we've moved. And so it's all by email or phone. So there's really nobody that I get together with now that can't see.

Team FACE Member: So would you wear it with or around sighted people?

Subject 8: Oh yeah. And I'd love to try it if it's a wrist thing.

11.1.9 Subject 9

Team FACE Member: So I have a few quick questions that are probably the easier ones that we can begin with. You don't have to answer any of these if you don't want to. They're just a demographic questionnaire. Can I ask your age?

Subject 9: 61

Team FACE Member: Ok. And your gender?

Subject 9: Male

Team FACE Member: And level of household income. You can also not answer this if you're not ok with that. We have from A through F. There are choices A) $0 - 19,999; B) 20,000 - 39,999; C) 40,000- 59,999; D) 60,000 - 79,999; E) 80,000 - 99,999; and F) over 100,000 a year

Subject 9: D.

Team FACE Member: How about ethnicity? We also have A through E in terms of choices A) American Indian/ Alaskan Native; B) Black or African American; C) Asian or Pacific Islander; D) Hispanic/ Latino; E) White; F) Other or Not reported

Subject 9: E.

Team FACE Member: And level of education?

Subject 9: Two years of college.

Team FACE Member: First question I have here is how do you describe your social interactions with other blind people?

Subject 9: Once the person who is blind is confident in themselves in their abilities to communicate with other people, it's fine.

Team FACE Member: What about with other sighted people?

Subject 9: Same thing.

Team FACE Member: Do you ever think there's anything missing when interacting with sighted people?

Subject 9: Not really. You can feel their voice and body flow. For instance, you and I are talking on the phone right now and facial expressions doesn't mean a thing to me and doesn't mean a thing to you. You hear it in a person's voice. I don't say I'm on the phone on all the time talking to people. But, I don't have to see their facial expressions to feel the warmth. You can I guess, but go on.

Team FACE Member: Do you currently use any assistive technologies or devices that help you in your social interactions?

Subject 9: No, not really. Nope.

Team FACE Member: Using a scale of 1-5, 5 being the highest, rate how important it is for you to recognize emotional cues in your daily social interactions - face-to-face.

Subject 9: Probably a 1 or a 2. Cause I've been doing it for such a long time. I haven't needed it in such a long time that it doesn't matter. I guess I'm hearing their voice, hearing how people talk. So I can recognize emotional cues that way too.

Team FACE Member: Rate your interest, on a scale of 1-5, 5 being the highest, in a portable device that could recognize faces and convey identities to you.

Subject 9: Probably a 1 or a 2.

Team FACE Member: Do you mind me asking if you were blind at birth or if you became blind later?

Subject 9: I became blind in my late 20s.

Team FACE Member: Would you rather have a device that would let you help you recognize who people are or give you feedback on expressions.

Subject 9: Give feedback on expressions

Team FACE Member: Is their a reason for that?

Subject 9: It doesn't really matter to me what a person looks like or who the person is. I try to portray that in some speeches I do the fact that I'm very lucky to be blind because I can't judge you on the color of your skin, on your height, on your weight, from the clothes that you wear, and its just from my heart to your heart. So that's really what it is. You may say look at that guy over there; he's gorgeous. And he may be the biggest jerk in the world. So visual world is not as important. You get a lot of feedback from vision, but you get a lot of feedback from being blind too. You just gotta use the things you have to do the things that you need to do.

Team FACE Member: What would be the best way for a device to convey facial expression cues to you, for example, via haptic (vibration) feedback, audio, beeps, chimes, etc.?

Subject 9: I guess if there was a device, I guess telling me smiling and frowning would be most useful compared to beeps and bumps.

Team FACE Member: Okay, so you think audio is best right?

Subject 9: Yup.

Team FACE Member: Do you think the audio would distract you from normal conversation?

Subject 9: Nope. For me, nope.

Team FACE Member: How would you rate our device?

Subject 9: I don't know. I'm not sure I would want cameras on my cane.

Team FACE Member: Do you think you would use it more if we made the device smaller?

Subject 9: Probably Yes.

Team FACE Member: Would you want the option of turning voice on or off during a conversation?

Subject 9: Yes, defintely.

Team FACE Member: And we have the last question: If you had such a device, would you wear it when interacting with other visually impaired people?

Subject 9: I don't know, I couldn't tell you that. You could make it so small and put in a cellphone and that would work.

11.2 Experiment Transcripts

11.2.1 Subject 1

Kailin Hsu: Alright, so just for our records, this is (Subject 1) and this is October 7th, 2011. Okay, so for the first experiment, we'll be playing a video for you and we'll ask you to identify when you think the speaker in the video is smiling, using a clicker. So you won't be using anything...

Subject 1: Okay.

Kailin: No device. It's just you, your own hearing - so no assistive devices. So every time you think that the individual in the video is smiling, just press the clicker. Okay, and do you have the clicker?

Subject 1: Where's the clicker?

Kailin: We'll give it to you in one second.

Subject 1: Oh, that's a computer, ok.

Lydia Lei: Here's the clicker. So the button is...

Subject 1: There we go and the button is there?

Lydia: Nope, on the top. So it's right here, where your thumb is.

Subject 1: Oh.

Hayato Unno: Yup, so that's what...

Subject 1: Oh.

Kailin: Ok, so can we reset it?

Hayato: Now we need to reset it.

Lydia: Yup I'm going to reset it.

Subject 1: I'm sorry.

Kailin: No, no, no. It's ok, we wanted you to test it out.

Subject 1: I wanted to see...

Lydia: Did you want to try it more?

Subject 1: No, I know how it works

Kailin: Do you have any questions regarding this experiment?

Subject 1: No.

Kailin: Ok, so just click when you think there's a smile.

Subject 1: Ok.

Lydia: I'm going to put your thumb on top...

Subject 1: ∗ singing ∗ Smiling faces, sometimes they don't tell the truth.

Lydia: So your thumb is on top of the clicker.

Subject 1: Yeah, got it.

Kailin: Hayato's going to give you a really brief intro to what the video's saying.

Hayato: So you're going to be listening to a video clip of an interview. The person being interviewed is actually a natural photographer. He likes to take pictures of animals in the wild. So, that's just a brief, I guess, intro to what you're going to be listening to.

Subject 1: Oh, ok.

Hayato: Yeah, so you're only interested in trying to figure out whether the person who is being interviewed is smiling.

Subject 1: Oh boy! Ok.

Hayato: Don't worry about the interviewer.

Lydia: Yeah, don't worry about the guy asking the questions, just the guy...

Kailin: Alright, are we ready?

Subject 1: Yeah.

Kailin: Alright, we'll play the video now.

(Video 1)

Lydia: Can you hear it? Is the volume too soft?

Subject 1: Hmm.

Lydia: Let's turn it up a little bit.

Hayato: Yeah, we should adjust...

Subject 1: Mmmm.

Hayato: We're just adjusting the volume right now. Is that fine?

Lydia: Is this ok? This volume level?

Subject 1: I think.

Lydia: Okay, let me start over then.

Lydia: Let's try plugging it in to the computer.

Subject 1: Oh, there you go.

Lydia: Let me see if this helps.

Subject 1: I wonder if it's on the...

Hayato: I'm sorry, we're having a little technical difficulty.

Subject 1: Okay. Due to technical difficulties...

Kailin: Was that the max volume on the laptop?

Lydia: Yeah.

Kailin: Okay.

Subject 1: Okay.

Lydia: We can just bring it closer.

Subject 1: There you go.

Kailin: Is this good for you?

Subject 1: It's better.

(Video 1)

Kailin: Ok. Great, so that was the first video clip .

Subject 1: Okay.

Kailin: And for consistency purposes, we'll be playing another, different, video.

Subject 1: Alright.

Kailin: And we'll ask you to do the same thing as before.

Subject 1: Same thing, okay, sure.

Kailin: Are you ready?

Subject 1: Yeah.

Kailin: Ok.

(Video 2)

Kailin: Alright, so that's the end of video two.

Subject 1: Okay.

Kailin: So...

Lydia: Can I take the clicker from you?

Subject 1: Sure, uh huh.

Kailin: Ok.

Subject 1: That's neat.

Kailin: So, the purpose of those two, er, that first experiment was to see how many smiles you could pick up without using any assistive devices.

Subject 1: Right.

Kailin: And then to compare how you felt during this experiment, the experiment that you just performed to how you will feel in the next experiment where we'll be giving you a bluetooth to wear in your ear.

Subject 1: Ok.

Kailin: Ok?

Subject 1: Yes.

Kailin: So this brings us to the second experiment. Do you guys have the bluetooth? Ok. We'll be playing two more videos for you.

Subject 1: Ok.

Kailin: But this time, they'll be two separate videos.

Subject 1: Alright.

Kailin: And we'll be giving you your bluetooth device.

Subject 1: Ok.

Kailin: So, the bluetooth will tell you when the person is smiling.

Subject 1: Hmmm.

Kailin: Right now...∗ to team members ∗ Are we going to play samples of the sound or no? We don't have to. Ok..so...alright... Do you want to give (Subject 1) this bluetooth? ∗ to Subject 1∗ We'll show you how to use it.

Subject 1: Hmm.

Alvin Hua: So this is the bluetooth. So you put that over your ear, that flimsy part that you feel right now.

Subject 1: Oh okay, yeah.

Alvin: Yeah.

Kailin: Actually, (Subject 1), I just found out that the chimes are actually embedded within the video. So they're actually not transferred through the bluetooth yet. So you won't need it for this.

Subject 1: Oh, okay.

Kailin: But, ideally, when our system is finished...

Subject 1: Yes.

Kailin: You will be using the bluetooth. And the sounds that you hear through the video will be transferred to you through the bluetooth.

Subject 1: Okay.

Kailin: Okay.

Subject 1: Alright.

Kailin: Sorry about that confusion.

Subject 1: No problem.

Kailin: So, for this experiment, we just...we have one video where you'll hear chimes corresponding to an expression.

Subject 1: Oh, ok.

Kailin: So, in our system, we have six expressions. So, we have smiling, frowning, surprise, disgust, anger, and...what's the...

Jia Zaidi: Fifth one? Fear.

Kailin: And fear. But, for the purposes of this experiment, because we have a time constraint, we're only going to be testing smiling. And because there's a learning curve...

Subject 1: Ok.

Kailin: ...For you to see which chime goes with which expression.

Subject 1: Mmmm.

Kailin: So, the chimes you hear in the video are corresponding to the times when the interviewee is smiling.

Subject 1: Oh, ok.

Kailin: Ok? So, for these experiments, you will not need to do anything when the videos are playing. We'll just ask you a series of follow up questions.

Subject 1: Alright.

Kailin: To gather feedback on our system.

Subject 1: Sure.

Kailin: Ok, so do you have any questions?

Subject 1: No.

Kailin: Okay, we'll play the first video.

Subject 1: This is fun!

Lydia: So this is with the chimes.

Subject 1: Alright. Hmm.

(Video with chimes)

Kailin: Okay. So that was the first video.

Subject 1: Alright.

Kailin: And those were the chimes.

Subject 1: Yes.

Kailin: So, in the second video, instead of hearing chimes when the interviewee is smiling, you're going to hear the system actually say, "'Smile."'

Subject 1: Alright.

Kailin: Okay, so, after this experiment, we're just going to ask you how you liked the feedback systems.

Subject 1: Sure.

Kailin: Okay?

Subject 1: Okay.

Kailin: So, start playing the second video.

Subject 1: Mmhmm.

(Video with "'Smile"')

Kailin: Okay, so that was the end of the second video. So, now we're going to begin the post-experiment questions.

Subject 1: Okay.

Kailin: So, on a scale from 1 5, 5 being the highest, like 5 being the most obtrusive, rate how obtrusive you found our feedback system. So you can talk about, so on a scale from 1 5, how obtrusive did you find our chime feedback?

Subject 1: Um 1.

Kailin: 1? Okay, so it wasn't obtrusive to you at all.

Subject 1: No.

Kailin: What about the second one? Where it actually said smile?

Subject 1: 2

Kailin: Okay, so you would prefer the chimes?

Subject 1: Probably, if I learned what they were. But you know, I kinda, its hard to say. Cause either one works, so I'm not...either way, really.

Kailin: Okay, does the feedback give you new information that you couldn't otherwise conceived?

Subject 1: Yeah, definitely.

Kailin: Okay, like what types of information?

Subject 1: Well it just tell me that the person, that there is a face, and that he's smiling or whatever the expression. And, I wouldn't necessarily know that otherwise,

Kailin: Okay, great.

Subject 1: Unless they were laughing or something.

Kailin: Right, And those gives you clues about those expression?

Subject 1: Yeah, or just the conversation. But say there was no conversation, there would be no way I'd know, you know?

Kailin: On a scale from 1-5, 5 being the highest again, rate how satisified youa re with the feedback system.

Subject 1: 5 being the highest? How satisified? 5.

Kailin: 5? Great, do you have any improvements, or suggestions for improvements that we could make to the feedback system?

Subject 1: Not that I can think of.

Kailin: Great, that's awesome. And this is kind of related to the other questions, but are there any other modifications, not necessarily improvements, that we can make to our feedback system to better suite your needs?

Subject 1: I can't think of any.

Kailin: Okay, so it's basically just the learning curve that's more of a concern for the chime feedback system?

Subject 1: Yeah.

Kailin: Okay, Do you have any concerns for the feedback system that actually says the expression?

Subject 1: No.

Kailin: So it's just the increase in obtrusiveness? The slight increase in obtrusiveness?

Subject 1: Mm, I mean. Either one is good, and it just depends on, I guess, preference, I would think.

Kailin: Do you think hearing the feedback would distract you in a conversation?

Subject 1: I don't know.

Kailin: Like a one-on-one conversation?

Subject 1: I'm not sure on that one.

Kailin: Okay, actually that's what we'll test later on, so we'll get your feedback on that.

Subject 1: Alright.

Kailin: So for experiment 3, we'd like you to play around with the device first. So, basically for this experiment, when the device sees a new person, you can ask it to record and remember that person for the future.

Subject 1: Hah.

Kailin: So yeah, these are ideally good for your friends, good for acquaintances you meet. But the system won't work on any random person because you don't have their information stored in your current database.

Subject 1: Oh, ok.

Kailin: We've already programmed the device to recognize three team face members, so three of us, who will stand in front of you. So someone's going to stand, we're all going to stand in a row in front of you, someone to the left, someone directly in front, and someone to the right. And we'll ask you to use the device to locate where we are. And then we're going to shuffle and see if you can identy our new positions.

Subject 1: Cool, alright. Hey this is cool!

Kailin: So right now, we're going to give you the device. So because this white cane is a little tall, we're going to ask if you can stand up.

Subject 1: Oh yeah, sure. If you can sit this up against the wall or something.

Hayato: So if you want to feel around a little bit. Our camera is on top.

Hayato: So right here, there's some, do you feel that bump right there?

Kailin: So the bump is to help you, because the lense is exposed in our camera.

Subject 1: Oh, so its like a focal point.

Hayalto: It's where the front of the camera is pointing. So if you want, the bigger bump indicates the back of the cane.

Subject 1: Oh ok, that makes sense.

Kailin: So the orientation makes sense to you?

Kailin: Yeah.

Subject 1: Great.

Hayato: It's not connected yet. So now we're going to give the Bluetooth back.

Subject 1: Oh okay. There we go. Yeah, I got it.

Hayato: Great. And now we're going to attach the camera to the computer. So there's going to be a cord like coming out of the back, so...

Subject 1: That's going to be interesting. How would they work something, where eventually, to where you're out somewhere and you don't have a computer? That would be interesting.

Alvin: Probably off of your phone.

Hayato: So that's something we're still working on.

Subject 1: I'm just curious.

Hayato: Right now, you have it pointing backwards.

Subject 1: Oh, wait a minute. Oh, there we go.

Hayato: So now its pointing forward.

Subject 1: It is?

Kailin: Pointing forwards in relation to where Hayato is, and Jia. They're standing right in front of you now and the camera is pointing directly at their faces. Just to give you a sense of where it's pointing.

Hayato: We apologize that the cane is kind of tall.

Subject 1: Oh well, hey, I understand. That was just kind of confusing because to me, that would be this way. Pointing this way.

Jia: It is pointing that way for now.

Hayato: Right now it's pointing to your left. That's correct. Its pointing at me

Subject 1: Right, that's what I thought. Alrighty.

Jia; Now its pointing at me, standing right in front of you

Hayato: So it's actually not on yet.

Subject 1: Alrighty. This is cool

Kailin: Its still booting up.

Alvin: Can we fix the Bluetooth for you? Its not quite on right.

Subject 1: Oh, sorry.

Kailin: Oh no, it's okay.

Subject 1: Oh, oh okay. Don't know. Just trying to do...

Hayato: So, Alvin is our technical person today.

Alvin: So, I'm going to start the system, you should hear a welcome. There you go.

Kailin: So right now, we have sound coming out of the speakers, like it was in the first two experiments. But right now we're trying to transfer the sound through the Bluetooth.

Subject 1: Yeah, to start out, we'll have to turn it down first. And then adjust it. Cool.

Lydia: What kind of computer do you have at home, (Subject 1)?

Subject 1: PC. Desktop. There we go.

Hayato: Is that too loud?

Alvin: No, is this too loud?

Subject 1: There we go. That's fine. Cool.

Hayato: Except we can't hear it. Should we use speaker?

Kailin: Well, we can try it through Bluetooth first. Because right now, (Subject 1) is just recognizing faces, right? Okay, so do you have the face recognition stuff set up?

Subject 1: Okay, it says, "'Welcome."'

Alvin: We can't hear it.

Kailin: Okay, so we're going to run a couple of tests first just to see if the camera recognizes these faces. So right now, it's not really an experiment, it's just for you to play around with the device to get yourself more familiar with it. So try pointing, not yet. Okay, wait one more second. The system is still loading.

Alvin: Yep.

Kailin: Ok. So try, Hayato's standing, ok, if you turn a little to...there you go. Hayato's standing right in front of you, so try to position...ok, a little more to your left.

Hayato: Hi.

Kailin: Ok, there you go, so it should be picking up Hayato's face soon.

Alvin: So to first start it, you must say, "'Initialize baseline."'

Kailin: So, (Subject 1) you have to say, "'Initialize baseline."'

Alvin: When you hear it play back to you, then you know that... did you hear it play back to you?

Subject 1: It said, "'Initialize baseline."'

Alvin: Alright good, now to try responding...ok...it lost you. Alright, now try.

Jia: Does it make a sound?

Subject 1: "'Who was recognized? No one. Smile."'

Kailin: Oh, so it said, "'Smile?"' Ok so that's the expression part.

Alvin: Um, right now our face recognition is not really working.

Subject 1: Oh ok.

Alvin: So you're going to get a lot of "'no ones."'

Subject 1: Yeah. Hi no one!

Kailin: But ideally, it should be able to identify someone and say that person's name, and like if you...so Hayato smile. Do you hear that?

Subject 1: Not yet.

Alvin: Try pointing away for a sec. Ok, now go back.

Kailin: You heard it the first time right?

Subject 1: I did hear it the first time.

Kailin: Ok. It's not supposed to be like this.

Subject 1: I hear ya. But that's why, that's why we're doing this.

Kailin: Yeah.

Subject 1: That's exactly why we're doing this.

Alvin: Ok, say who is recognized.

Subject 1: I don't know. Somebody.

Alvin: What did it say?

Subject 1: It just said, "'Who was recognized?"'

Alvin: And did it say anything after that?

Kailin: Did it say anything after that?

Subject 1: What is the current expression?

Alvin: Who was recognized?

Subject 1: It said, "'Who was recognized?"' and "'What is the current expression?"'

Kailin: Alright.

Subject 1: Oh, something, "'Hioto,"' or something.

Kailin: Hayato?

Hayato: Oh. How did it say my name?

Kailin: Great, so it identified Hayato.

Subject 1: Oh cool.

Hayato: It said my name wrong. but that's ok.

Kailin: Alright, so let's try Jia. Are you on the system? Ok so try directing your camera towards Jia. Jia talk.

Jia: I'm right here

Kailin: There you go.

Alvin: A little to the left.

Jia: There we go.

Alvin: It's not picking up her face yet. There you go.

Lydia: Do you hear anything (Subject 1)?

Subject 1: Not yet.

Alvin: Say, "'Who was recognized?"'

Jia: Go ahead and ask the system who was recognized and it will respond to you. So you can try.

Subject 1: Who was recognized? Jia.

All: Nice.

Kailin: So you actually have to talk to the system to prompt it.

Subject 1: Oh I didn't know.

Kailin: Sorry, we didn't give very clear instructions about that.

Alvin: But in the future, you will not have to talk to the system. It will tell you when it recognizes a face.

Subject 1: That's almost like this GPS system that you've got too.

Alvin: Right.

Kailin: Yea, ok. so should we try some....

Subject 1: Let's try expression. Smile.

All: Nice.

Kailin: Right, she's smiling. Ok, should we test someone else or go directly into the experiment?

Hayato: Let's play around some more but maybe this time...

Jia: Well, does it show angry yet?

Kailin: I don't know.

Jia: Try making an angry face.

Kailin: We're going to try making different faces. So you're just going to ask the system, "'What is the current expression?"' and the system will tell you.

Subject 1: I don't know when I'm supposed to do that.

Jia: You can ask now.

Lydia: Whenever.

Subject 1: What is the current expression? Who was recognized? Surprise.

All: Yup, there we go.

Jia: You can ask again.

Subject 1: Who was recognized? No one. What is the current expression?

Lydia: Did it say anything?

Subject 1: What is the current expression?

Alvin: Actually, that doesn't do anything.

Subject 1: Surprise.

Alvin: It's just when the expression changes. It'll tell you automatically.

Jia: Oh it tells. Ok.

Subject 1: Surprise.

Lydia: So it keeps changing.

Alvin: So it keeps changing. You don't have to ask the expression.

Subject 1: Smile. Smile.

Jia: I tried surprise. Ok, I'll try again.

Subject 1: Smile.

Kailin: Open your mouth really wide.

Lydia: Raise your eyebrows.

Subject 1: Surprise.

All: There you go.

Kailin: Ok. Can you try frown?

Lydia: Try angry.

Jia: Not really. I've been trying angry.

Alvin: Anybody's expression can be recognized, just not your face.

JIa: You try. You have a better angry face. So Hayato's going to try and make a angry face.

Hayato: I don't really have a good angry face.

Alvin: That's ok.

Subject 1: Who was recognized? No one. What is his expression?

Kailin: It's supposed to be angry.

Jia: Maybe try changing from neutral to angry?

Alvin: It's saying you're neutral right now. And now it's saying your surprised and neutral.

Hayato: How about a smile?

Alvin: Smile. Actually now it's saying you're disgusted.

Kailin: Do you hear all the feedback?

Subject 1: No.

Kailin: Ok.

Hayato: Would it be easier if we played the sound on speaker so everybody can hear?

Kailin: Ok, so right now we're just going to switch off the bluetooth and put it on speakers so that we're able to hear what the system is saying to you.

Alvin: But you should keep the bluetooth on so you can give commands.

Kailin: Is that really loud? Alright, Alvin's just working on transferring the sound.

Alvin: So, initialize baseline.

Subject 1: Intitialize baseline.

System: Initialize baseline.

Alvin: So you can say things...so to start the expression and facial recognition you have to say initialize baseline.

Subject 1: Initialize baseline.

System: Intialize baseline.

Alvin: You can also say as you know, what is the current expression?

Subject 1: What is the current expression?

Alvin: And who was recognized?

System: What is the current expression? Smile.

Subject 1: Ok. Who was...?

Alvin: Who was recognized.

Subject 1: Who was recognized?

System: Who was recognized? No one.

Alvin: Right. You can also change the expression feedback mode. So you can either use chimes or words. And we're gonna add another one called onomatopoeia.

Subject 1: Oh.

Alvin: And that's just like, when you smile it's going to say, "'yay."' So you can say that command: "'I want to change the expression feedback mode."'

Subject 1: Change the expression feedback mode.

Alvin: I think you have to say, "'I want to..."'

Subject 1: I want to change the expression feedback mode.

System: I want to change the expression feedback mode. Change to words or chimes?

Subject 1: Chimes.

Alvin: For that you have to say, "'Change mode to chimes."'

Subject 1: Change mode to chimes.

System: Change mode to chimes. Mode changed to chimes.

Alvin: And now you should hear that chime you heard earlier whenever he smiles and Jia smiles.

Hayato: I'm standing right here and Jia's standing right in front of you.

Jia: Over here.

Subject 1: Oh, yup. Got it.

Alvin: Yup.

Jia: I did.

Lydia: Oh.

Alvin: Ok, and that's all we really have right now.

Subject 1: Ok.

Kailin: Ok, so I guess our face recognition system isn't working? Ok, so I'm sorry but we won't be able to..huh?

Jia: It is on me so she could try locating me.

Kailin: Ok.

Jia: I could be one of the three people.

Jia: So we have 3 positions in front of you, set up for us to be standing in, to the front, left, and center. Now you know how to ask system to ask person...Now you can try alternating between three positions and find where I am. Someone else will tell you we have changed positions and you will be able to use the device.

Kailin: You have to move the camera relatively slowly so that you will have a chance to pick up Jia's face.

Jia: For the first time, I am right in front of you, Lydia is on the right, Hayato is on the left

Alvin: After you initialize baseline, the system running. Ideally, when a face is up, you will be able to tell who is recognized. But right now I will tell you which is recognized.

Subject 1: Which face is recognized?

Jia: We will switch, and Alvin will tell you.

Alvin: Try pointing to the right...a little bit, a little to the left. Stop right now. Can you move a little closer?

Kailin: Okay, it's not picking the face up.

Alvin: You can ask it now

Subject 1: Which face is recognized? "'Which face is recognized? Hayato."'

Kailin: Hayato is standing directly in front of you, can you find where Jia is standing?

Alvin: You want to ask it, "'Who was recognized?"'...Hayato...that is a fault of our system....go to the right a little bit.

Subject 1: Which face is recognized?

Kailin: Ideally, whenever the system identifies a face, the system will tell you who was recognized. After that the system will tell you who the actual person was.

Alvin: Say, "'Initialize baseline,"' one more time

Subject 1: Initialize baseline. Who was recognized?

Alvin: If you say, "'Initialize baseline,"' it will start recognition. A repeat cancels recognition. Say it again. Start again.

Subject 1: Initialize baseline. Who was recognized? "'No one."'

Kailin: What do you think was the least useful part of our device? If the facial recognition part was working.

Subject 1: I don't know. I guess the least useful...hmm...'Cause you know what I think, I think it's kind of neat. I think if you had it set to the two voices, I mean, you know, like the recognizing of the face and expression that might be distracting. I think it's better to use it with, you know, the tones for the expression and the identifying of the face with words, that's what I think.

Kailin: That's really good feedback.

Subject 1: I think that would be wisest.

Kailin: Ok, yeah, that's great. Do you...Are there any modifications you would make to our device to better suit your needs besides the one you just mentioned?

Subject 1: Well, I'm thinking...See my thing is sometimes..is...You know I'm not sure, is this...if this is supposed to be in a room where I'm identifying a face, that's one thing. If I'm out on a city street and I meet someone, usually I guess they would identify themselves, "'hello"' or whatever but suppose it's somebody I don't know and they're trying to be helpful, what I'm thinking is, "'What is

their expression? And are they really going to be useful or are they going to be damaging?"' I'm thinking of the fact that really, you got to kinda get a sense of, you know, is this person safe or should I steer clear of them. Yeah, that's self-defense. If you're going to use your cane, that could damage the equipment.

Kailin: Yeah, that's true.

Subject 1: So I'm like, you know...Because number one, the cane is going to be shorter so how am I going to aim for a higher face?

Kailin: Right, that could be used when you're sitting down.

Subject 1: Yeah, see that would be...I would have to keep my cane the way I have it unfolded, you know. Or I could fold it and kind of move it around. You know, that would work.

Kailin: Ok, so there are there canes that that don't fold and there are canes that fold?

Subject 1: Right.

Kailin: Ok, give me one second. So...

Subject 1: So the thing is when you are going to put this thing away, this contraption away, it's got to be enough that you can put it in your pocket or you can put it somewhere where it's not going to be damaged if you fold the cane. You know, that's another thing - the durability.

Jia: Well if you like, you can look - the camera actually screws off the top. So, this is just the camera, so you'd be able to remove it and that would fit into your pocket. So this is just the camera.

Subject 1: I gotcha.

Jia: And you can even put it onto a tripod so if you're sitting down and you don't want to have your cane standing straight next to you, you can always place

it onto the table or screw it onto a tripod.

Subject 1: Very seldom do I ever fold my cane, but when I do, it's in a situation where I'm not too concerned about who's what and whatever...Oh, this is neat!

Jia: That's a protective case for the top. So if you accidentally drop the cane, it won't exactly hurt the camera as much if it wasn't protected.

Kailin: (Subject 1)...

Subject 1: How does this deal in the weather situation?

Jia: Well it's made of plastic, it has vents on the side. This is the camera.

Subject 1: Yea, it's warm, uh huh.

Jia: And this, the thing you are holding in your hand, that's plastic so it covers it and protects it from water, protects it from damage, and it also has vents on the side .

Subject 1: Yes.

Jia: So it keeps it from heating up too much.

Subject 1: I gotcha.

Jia: So we've tried to make it weather friendly.

Subject 1: Yeah.

Jia: Uh, accident friendly.

Subject 1: Yeah, that's what I'm thinking about. Just see because, you know, sometimes I have fallen and I have crashed and one time I broke my wrist and had the cane in my hand and that's the one I broke and, and so that could have been dangerous to that poor camera. It could mess up something I had on that cane that it was used for lighting you know having light when I was out on the street. You know, it had a flashlight thing that got damaged so, I mean, like,

and I never got another one of those again so...Because, see, and the other thing is see the weight is going to affect a person's movement around, so if they're used to a lighter cane and then the weight is put on there, they need to get used to walking around with it.

Jia: Did you feel it was substantially heavier than canes you're used to using?

Subject 1: Well see, I didn't try to travel with that cane so there was no way I could've figured it out and since it's attached to a computer, you know. But I would think that it might affect...so you'd have to get used to it, you'd really have to get used to walking around with something like this. I'm wondering if there was any other way to maybe mount something um that might not necessarily be on your cane and you're walking...I mean it might be a good thing to be on a cane, but I'm kind of thinking...I'm just wondering if it would be versatile enough if you could mount it, say...Of course you don't want people thinking "'Oh she's got a special piece of equipment, let me steal it."'

Lydia: Right.

Subject 1: So that might not be a good thing, so I don't know, you know, 'cause you're dealing with walking around out in traffic. Gotta pay attention to traffic.

Jia: Right.

Subject 1: You gotta pay attention to...Even in a building, you gotta pay attention if you're walking around. You gotta pay attention to where things are, elevators and various things and if you're in a crowd, you gotta know what you're doing. So I'm looking at all sorts of situations where you would carry this. If it's only designed for being in a quieter place than in a room, then that needs to be stated. If it's designed also for travel that needs to be stated. If it's designed for

a room thing than you wouldn't need to necessarily have it on a cane, so that's where I'm coming from.

Lydia: So if we specified those different uses, would you use it in a real life setting?

Subject 1: Yeah, I mean, but you'd have to...but how do you use it in situations, various situations, or is it only designed for a certain set of situations? That's where I'm coming from.

Lydia: That's a good point.

Jia: Well, our...When we initially designed it, what we were trying to focus on was, let's say you walk into a party and you want to know who's standing in front of you. You want to ask the device to tell you whether they're happy with the conversation, so if they're smiling. Then also, in a meeting, we thought maybe ...

Subject 1: Well, in a party, put it to you this way, you've got a number of people in there. I've been to alumni things, let me put it that way, you know how loud it gets in there? You got people shouting, because you got hundreds of people in there. You know the volume gets so loud as far as people talking and everything. 'Cause, I know I was at this one and there were like three hundred people in there and it was a huge gathering and it was like we were loud. 'Cause I mean it was so loud, I had to shout to talk to somebody. So its like, it was kinda neat, but people had to identify themselves. And then we would talk and everything but it was kinda like in some situations its like a different thing.

Lydia: Right so this really wouldn't work in a party setting, I guess?

Subject 1: No.

Lydia: Or if you're not even able to hear the person you're talking to, I guess.

Subject 1: Or if you hear them, but it's like you've got everyone else's voices too. So you've got all these people talking. So you got to pay attention to them and pay attention to this too. I'm not sure that would [work]...In a meeting situation, that's one thing. But in something like that...Thats why I question about the travel too. If you're in a quiet country area, or a rural area, or a quiet area of a city, or in a building, you might be okay. But I do a lot of traveling in cities. I travel on the trains. The trains are underground a lot of times. Where is this (the device) going to work? So you know, there are a lot of questions.

Lydia: Ok. Do you have any other feedback for us?

Subject 1: I can't think of any.

Lydia: Ok, that's fine. Kailin went to go put more money in the meter, so I'm going to take over for the last experiment. We're about to start experiment number 4. To make you more comfortable with the device in a natural setting, we're going to let you use the device during a one-on-one, face-to-face conversation. We'll have a relaxed conversation about various topics, and we'd like you to evaluate our system while we're speaking and give us any feedback you may have about our device. And we will be using chimes in this experiment to indicate expressions instead of words.Ok, so you're going to have this conversation with Jia, who is sitting to your left. So you guys can go ahead and face each other. We're going to point the camera at Jia so that the camera will be analyzing Jia's expressions.

Subject 1: Ok. That sounds neat.

Hayato: This time, the camera will be mounted on a tripod.

Subject 1: So it's almost like I'm interviewing somebody, like for a job.

Lydia: Yea, correct. Something like that.

Subject 1: Now that would be an ideal thing. You know, say you're interviewing somebody.

Jia: Right. And since you have a Bluetooth, you'd be able to hear feedback about whether or not they're smiling, if they seem enthusiastic. And they won't know that you're actually able to see all of this.

Subject 1: Aha! You got it.

Lydia: (Subject 1), are you comfortable? Would you like us to turn your chair a little bit?

Subject 1: Alright. There we go.

Alvin: Ok. Let's set up a little bit. So you will have chime outputs.

Subject 1: Sounds good to me. Let's roll.

Hayato: This is really cool. We never thought of using our device in this kind of situation - in like an interview.

Subject 1: Yeah, I'm thinking job interview. Like talking to a lawyer, or you know...

Hayato: In a situation where you really want to get to know who you're talking to.

Alvin: Change mode to chimes

Subject 1: Change mode to chimes. (Prompts system to repeat commands) You know what would be ideal? If there was a button mode or something where you could change the modes, where you didn't have to pump in all those commands, so that nobody would know that I was carrying [the device]. Otherwise, I'd have to explain, "'I have this piece of equipment, and it's supposed to monitor people's faces."' I don't want to do that because they might be afraid to have an expression go a certain way. And some people don't like being recorded.

Lydia: Right. So the good thing about our device is that it doesn't actually record video, but it just has a live stream of video.

Subject 1: No. But what I'm trying to say is that if they know that you're watching their face.

Jia: I think it is picking up words from our conversation, and taking them to be commands.

Subject 1: I'm sorry.

Alvin: It's not your fault. Change mode to chimes. (System continues to repeat commands in the background). Initialize baseline.

Lydia: So you guys can begin your one-on-one conversation.

Jia: So you can ask me to smile, so you can hear the chimes. Just to review.

Subject 1: Smile.

Jia: Now you can ask it what expression it shows.

(system doesn't work)

Subject 1: Hmmm. See, that's not useful.

(Jia and Alvin try to fix the system)

Subject 1: What I'm trying to say is, if you're in an interview, you don't want to be talking to the machine and saying, "'What expression?"' or "'Who's the person?"'

Jia: Well, when you're in a conversation, what we're also trying to make our device do is to have it continuously give you feedback like it did in the second experiment. As you were listening to the clip, it would also say, "'smile,"' or make the chime noise whenever the smile came up without you having to ask it.

Subject 1: Let's say I didn't know a person, and the person introduces himself. That would automatically feed into the system. See what I'm trying to say?

Instead of you trying to worry about setting all of this up, with the baseline, just [have the system] pick it up. I mean, I know I'm kind of strange, but that's what I'm thinking. So when the person introduces himself, and I introduce myself and I say, "'How do you do?"' and whatever, then it's picked up that that person is whoever that person's name is. And then we could go to the interview, but they don't know that that's what's going on. You don't want them to know. You don't want the other guy to know that you've got some kind of weird device. You really don't. I mean that's safety. You want to be as natural and as though you were a sighted person. You don't want the blindness issue to get in there. That's why I'm saying [your idea] is a great thing, but you've got to work it in such a way that nobody's going to know that I have a device until later or something. For example, I could go out into a group and they say, "'How do you recognize facial expressions?"' Or say you talk to a bunch of kids in a school and they say, "'Can you recognize facial expressions?"' And you say, "'Oh yeah, I got this device,"' and then I show them. But for normal circumstances, you don't want that. Because if somebody did attack you, then you could go to court. And you could say, "'Yeah, I know this person's face because I had this device. They didn't know that I had it."' And then it could be used for a lineup. And maybe I could tell if that was the person. See where I'm coming from? There are a lot of uses for this thing, but you don't want the other guy to know. Maybe that's a strange idea, but that's what I'm thinking. You don't want the other guy to know because then they'll try to make their expressions the way you want it.

Jia: Yes. You don't want them to give you the wrong kind of impression.

Subject 1: So if the person's not suited for whatever [position], you'll know. But they don't until you say, "'I'm sorry, but I don't think you would be suited

for this particular position."' Because I interview people for network marketing stuff. I'm into business that way. So it's like I introduce them to something, and I say, "'Take this card and check this site out."' Or I might say, "'Let's watch the video together,"' and then I would ask them what they liked about the video. I want to know honestly what their expression is, but I don't want them to know.

Jia: Right. You want the very genuine expression

Subject 1: Right. Am I making sense?

Jia: Of course you are.

Alvin: Initialize baseline.

Subject 1: I've got all the time in the world. I could work for hours on this.

Kailin: (Subject 1), can I ask you if you were born blind or if you lost your sight later in life?

Subject 1: I was a premie, so they thought it was oxygen that caused the blindness, but now they think it's the incubators and lights.

Jia: Do you know of anyone else with the same history, who said that maybe incubators could cause blindness?

Subject 1: Well, there are a number of people all over the place who have that kind of blindness. It's called retinopathy of prematurity, so I guess you would have to ask if they were born blind. Really, I don't know what it's like to see at all.

Jia: Because one thing we we're interested in finding was that if expression or recognition seemed more important to people who have congenital blindness or who became blind later in life.

Subject 1: I'm thinking that it might be either one, or the group who has seen before but, then again, who's to say? You have to ask. That'd be one of the

questions have you ever seen before and would this be beneficial more so to you or someone who smiled?

System: Smile, smile, smile...

Alvin: Works now.

Subject 1: Interviews...Let's say somebody is a blind lawyer, we have a lot of those, and they're doing deposition. Think about those court cases, and they sit there. They'll ask a certain question and later on they'll ask the same questions but framed in a different way to try and see if the person will give the same answer. And they keep doing it. I transcribed one one time and it was horrible. It was like they won't leave this person alone. They keep asking it. And this poor woman was like...It was a case dealing with the federal government versus a certain person (mentally ill). It was a case where they wanted to make it so that the mentally ill are in the least restrictive environment. And so it was like it was [] hospital versus...I was typing this deposition and it was just crazy. I worked overtime on a Saturday on this. I would never want to be in a court case for a lawsuit or some kind of problem. Maybe as a witness or jury I would love it. That would be another thing! If you're in a jury, of course you don't want them knowing you've got this machine because you can't have a cell phone or something going off but, what would be neat is that if you sit there and recognize that this person in this case, the defendant or something, has a certain expression when they're up there on the stand. Or the witness. You would be able to probably get an idea of if this person is guilty or not guilty. So consider that one. It would make us better able to serve on the jury, that would be cool.

11.2.2 Subject 2

Kailin: And just for our bookkeeping purposes, this is (Subject 2) and it is Friday, October 14th, at 2PM. We just went through introductions. Thank you again for participating in our experiments. Before we begin today, I just want to go over what our device consists of so that you know what you're using. There are 3 pieces to our device a Bluetooth to wear on your ear to obtain feedback (and give controls), a camera mounted on a white cane, and a laptop to power the system. The camera allows you to identify the people who are in front of you and the expressions they have on their faces. The Bluetooth will emit a sound when a person is in front of you and will convey certain expressions, like whether they're smiling or frowning. We will be asking you to participate in four experiments today. In our first two experiments, you won't be using the device, but you will be for the last two. We'll let you play around with the device during experiments three and four so you can become more familiar with it and its different features.

For the first experiment, we will be playing two videos for you. We will ask you to identify when you think the speaker is smiling. There's a pen in front of you, and we'll just ask you that whenever you think the speaker is smiling that you just click the pen. So the video is an interview. There is the interviewer and the interviewee. We would like you to focus only on the interviewee and when the interviewee is smiling. Do you have any questions?

Subject 2: No.

Hayato: The person being interviewed is a nature photographer. He likes to take pictures of wild animals. That's just the background.

Subject 2: What's the purpose of the pen clicking? Can't I just say, "'Oh, he's smiling."'?

Kailin: Just so that we can hear when you think he is smiling. Because if you say that you might miss another smile...

Subject 2: So is there going to be more than one smile?

Hayato: That's for you to decide. You can click as many times as you want.

(Video 1)

Carol: So that was the first video.

Subject 2: Ok. I thought he might have been at first, but it was sort of hard to tell.

Kailin: So should we put that down as possibly one?

Subject 2: He was possibly smiling when he was talking, but it wasn't clear-cut that he was smiling.

Kailin: Ok. For consistency purposes, we're going to play another video. We're going to ask you to do the same thing as before. Ok. Start video 2.

(Video 2)

Subject 2 (during video): He's still smiling (after hearing continuous laughter). They're both smiling (referring to interviewer and interviewee in video).

Carol: So that was video 2.

Subject 2: He has sort of a smiley voice anyway, doesn't he? He's sort of a happy person anyway, so I would think that other than the part where they were laughing, he's sort of a smiley person anyway.

Kailin: Ok. So you think he has a smile on his face the whole time pretty much?

Subject 2: He has a pleasant expression a lot of the time. He's probably sort of a smiley kind of person anyway.

Kailin: Ok. So that was the end of the first experiment. And the purpose of this experiment was to see how many smiles you could pick out without using any assistive devices, and then to compare how you felt during the last experiment to how you will feel in the next experiment where you will be hearing exactly when the interviewee is smiling. Ok. So for experiment 2, there are already sounds embedded in the video to indicate when the interviewee is smiling. So we'll be playing two more videos again. For the first video, you'll be hearing chimes that correspond to when the person is smiling. And in the second video, you'll be hearing a voice that says smile when the person is smiling. You don't need to do anything during this experiment.

Subject 2: I just find out how well I did.

Kailin: Exactly. Actually, these are two different videos.

Subject 2: Oh shoot. Haha, I wanted to hear the same one!

Kailin: Oh, and just to let you know. For our system, we are working with six expressions. But for the purposes of this experiment and due to our time constraint, we're only working on smiles. And because with chimes, there would be a larger learning curve.

Carol: Ok. So this is the video with the chimes.

(Video with chimes)

Subject 2: When it chimed like three separate times, were those three separate smiles? Do people do that?

Hayato: When people are talking, their lips are constantly moving. So what our system would do is that whenever the face forms the shape of a smile, it'll beep.

Subject 2: He could have just been saying a word that looked like a smile.

Hayato: That's possible. Yes.

Subject 2: Hmmm. Ok.

Kailin: So that was the chimes. In the next video, you'll hear the voice feedback.

(Video with "'Smile"')

Subect 2: It's funny, his smiles are coming in middle of phrases. What is the lag time between when he says smile because people don't normally smile in between phrases, do they?

Andrew: Sometimes they do, cause there is a slight pause when you talk. Sometimes you smile when you talk about something passionate

Subject 2: Normally, people start smiling before they start talking, don't you?

Rosie: During the video so he is actually smiling in between phrases. So the smiles that you are hearing, he is actually also smiling in the video.

Hayato: There is a lag time but it's pretty much a matter of miliseconds.

Subject 2: I didn't know people did that, huh. Okay.

Kailin: On a scale of 1-5, 5 being most obtrusive, rate how obtrusive our chime system, 1 being the least and 5 being the most obtrusive...and our voice system

Subject 2: 1 - I didn't find it obtrusive at all. I'm used to things like that. I'm used to things talking over each other. It doesn't bother me. I don't think you will find it to be obtrusive for a lot of blind people

Kailin: So you don't think it would interfere in your daily conversations at all?

Subject 2: It might a little bit. I would say 1 for the chimes and 2 for the voice.

Kailin: Does the feedback system give you new information you would not otherwise of conceived?

Subjct 2: Yeah, I didn't know people smiled in the middle of when they are talking. That was surprising to me.

Kailin: On a scale of 1-5, rate how satisfied you were with feedback system, 5 being the highest.

Subject 2: I personally think that you could probably get enough info about people's experssions by their voice. I think the facial recognition (I'm anxious to see that), for a lot of blind people, would be the most important thing. For the expressions, it certainly was what I expected, so I was certaintly satisfied. I would say 4, it did what you said it would do.

Kailin: Are there any improvements you suggest?

Subject 2: For expression recognition, no, but for those of us who don't use canes, the camera part would need to be attached to something else. Like on wrist, or wristband.

Kailin: We have another prototype where the camera would be connected to straps of a backpack.

Subject 2: Ok, so the laptop thing would be in the backpack. That would be fine.

Kailin: Out of curiosity, are you congentially blind?

Subject 2: Yes.

Kailin: So we are trying to see if one of the groups (congenitally versus non-congenitally) would have preference for expression or recognition, or vice versa.

Subject 2: I would think that maybe people who are blind later in life would find that they miss facial expressions. To me, it doesn't matter. A lot of times if

I'm lookiing for a particular person and they are coming down the hallway towards me, I want to know before they are gone. I would prefer for facial recognition. Facial expression is interesting, but I don't know it would be something that I would lug in a backpack. If I could get facial recognition, I would carry a hell of a big device for that.

Kailin: We will start with experiment 3 soon, but we will give you the opportunity to play with the device. Andrew will introduce you to the device.

Andrew: So a little disclaimer, the device isnt perfect. We still need to work on it. I am about to hand you the cane.

Subject 2: Oh god. I always lose those.

Kailin: There is an opening on the lens, so try to stay away from the lens opening.

Subject 2: How tall is this thing?

Andrew: It is really tall, I would say it's about um

Subject 2: This isn't really a blind person's.

Kailin: It is, we got it from NFB.

Subject 2: Oh god, that's the problem.

Andrew: Use the button to align the cane, it indicates where the front of the camera is. We also have a bluetooth device that you can wear on your ear.

Subject 2: Does it matter which ear?

Andrew: Let me see, Got it? Let me turn the device on. You will hear everything through the computer speakers for now.

Subject 2: What is the range does this thing have? How close does a person need to be?

Andrew: Right now, you have to be fairly close, but we are trying to work on

making that better.

Kailin: (Subject 2), where would you recommend getting a white cane?

Subject 2: Well, a place called Whiteplace and Moore or Ambutek and Revolutionary Enterprises. NFB uses canes that are way too long. That's ridiculous.

Andrew: To demo the device really quickly, Carol, can you start speaking? If you point device at Carol, try it. It was supposed to give you some feedback. Now if you ask the computer who was recognized, the computer should tell you, "'Carol was recognized."'

Subject 2: Do we just ask computer, "'Who was recognized?"'

System: Carol.

Subject 2: Ok, let's go! Who was recognized? ∗ repeatedly - testing on Carol ∗

System: Smile.

Kailin: But it's telling you that Carol's smiling at you.

Subject 2: How do you put people in? I'm more interested in facial recognition.

Andrew: So right now, we actually don't have an easy way to add people to the system. But what you would do would be to say something like, "'Save image for recognition,"' or "'for future recognition."' And then it would save that image into a folder.

Subject 2: So you would have to get the cooperation of that person?

Andrew: Right.

Subject 2: At first, and, "'Hey, can I take your picture so I can know who you are next time?"'

Andrew: Yeah, they would have to cooperate.

Subject 2: Yeah, that's fine. So then you would have a database of all the people you want it to recognize.

Andrew: Yeah, exactly.

Subject 2: I want one.

Andrw: So the problem we're having right now is going from speech to text. It's a lot easier for the computer...

Subject 2: Well, why can't there be a button on here that I could press instead of saying, "'Who was recognized?"' Wouldn't it be better just to have a button?

Kailin: That's a really good idea.

Subject 2: So you could get rid of the voice recognition. Especially if you're in a crowded neighborhood.

Andrew: Yeah, that's a great idea. Like I said, it's not perfect.

Subject 2: "'Who was recognized?"'

System: Smile.

Lydia: Well, at least we know expression is working. (tries different expressions)

Subject 2: Who was recognized? (repeatedly)

(system gives wrong answers)

Alvin: I think the bluetooth might not be working.

Subject 2: I went to do a presentation once with a wireless network and the wireless network wasn't working. So we talked a lot.

Lydia: What kinds of presentations do you do?

Subject 2: It was about the Library of Congress' program where you can download books. There's about 25,000 books on there now, so we were talking about how to do that.

Lydia: Is that available to the public?

Subject 2: No, only blind people. You have to be blind or handicapped. You can go to the library!

Lydia: Oh, that's true.

Alvin: Let's see if the recognition is working. So Lydia, can you...

Lydia: (Subject 2), do you want to direct that at me?

Alvin: Okay, it captured the face now, so, "'Who was recognized?"' Interesting. "'Who was recognized?"' No, it's not catching anything.

System: No one.

Andrew: We got the bluetooth working again. "'Who was recognized?"'

System: No one.

Andrew: "'Who was recognized?"'

System: Carol.

Subject 2: You know, it might be nice...my husband was suggesting it this morning - to add to the facial recognition - it might be nice to add a description. You know, is this person so and so. Is he or she tall, thin, well, thin's a subjective word, blonde hair, blue eyes. You know, it might be interesting

Lydia: Just to know.

Subject 2: Just to know, like so and so is a roly-poly. Really, the facial recognition is a big "'I want that."' Actually the police have that, there's an iPhone app. The app - it's like $ 3000. The police are using it to identify criminals in a criminal database or something. So if they can do it, we can have it too.

Andrew: Okay, so I'm going to add more people to the database.

Subject 2: Okay.

Andrew: And then, hopefully, the recognition will get better. Pretty much the only command that we really need is, "'Who was recognized?"'

Subject 2: And what it really needs to tell me to do is tell me when the focus is done. Especially if your laptop is in the backpack, you won't know. How is that going to work? Because laptops usually don't work when they're closed

Andrew: We can change the settings on the laptop.

Alvin: "'Who was recognized?"'

Andrew: Okay, add more people to the database

Lydia: (Subject 2), during this time, we actually have a couple other chime options that we'd like to play for you.

Subject 2: Okay.

Lydia: This is just a time for us to get feedback on what type of chime you prefer in our system.

Subject 2: Okay.

Lydia: It's now currently being set up.

Subject 2: Yeah the one that I heard before sounded a little like the chime on my GPS. That isn't my favorite sound. It's okay, it does its job. But I'd like a more pleasant sound.

Lydia: Do you actually want to know what sound that was?

Subject 2: Huh?

Lydia: That chime was actually my voice. Yeah, because we couldn't find a chime online so we decided to record one...

Subject 2: Oh there's plenty online, like Microsoft sounds - just get one of the sounds from Microsoft.

Lydia: That's actually...we were actually going to look into that.

Jia: Well, what we've been trying to do so far is come up with chimes that correspond to happiness, sadness, anger, and surprise, and make them as intuitive as possible.

Subject 2: Okay.

Jia: So that's where the difficulty was, finding chimes that were similar, yet different enough that you could pick out which expression they were for.

Subject 2: Right.

Jia: So we've been trying a few different instruments.

Subject 2: Okay.

Jia: We're looking into Microsoft sounds and we have some of those samples to show you.

Subject 2: Okay.

Carol: It's playing all of them at once! So, this is one possible sound.

(Sound 1)

Subject 2: Okay, certainly unobtrusive.

(Sound 2)

Subject 2: Oh that's a happy sound. I like that.

(More sounds)

Subject 2: See a lot of them are pretty similar.

(More sounds)

Subject 2: Maybe...there was one that was a definite happy one.

Kailin: That's the second one.

(Play sounds again)

Subject 2: Yes, that one is a good one for the smile. But maybe one's that go down in pitch like $*$ demo $*$ could be a frown or an unhappy $*$ demo $*$ could

be for angry, like a buzz noise ∗ demo ∗ could be for surprise. You know? You could create a sound....

Gems member: Yeah I think we tried.

Subject 2: Are they just wavefiles?

Jia: Yeah.

Lydia: Thank you for your feedback.

Subject 2: Sure, I give feedback all the time.

Andrew: I think we could probably start the next experiment now.

Kailin: Alright, so I'll explain experiment 3, who's participating?

Andrew: Hayato - we have pictures of Hayato. Hayato, Lydia, and Carol.

Kailin: Okay. So for experiment 3 we're going to be testing out the facial recognition system. And we've already programmed the device to recognize Hayato, Lydia, and Carol.

Carol. So we're going to ask these three Team FACE members to stand in front of you in a row. One will stand to the left, one will stand in the center, and one will stand to the right. And then we'll ask you to use the device to locate them as they go through a series of shifts. Okay, so they're all lined up right now, so we'll give you the device. And you can locate where team member is. Do you want to have (Subject 2) stand right in front of Carol because Carol's in the center right now?

(Side conversations to position (Subject 2), (Subject 2) jokingly directing a dog)

Jia: So we have three of them right in front...so for the first one we'll go ahead and test. So Carol's right in front of you.

Subject 2: Who was recognized?

System: "'Who was recognized?"' No one.

Alvin: It doesn't have a face yet. Try now.

Andrew: Well what we can do is tell you when the face is found because that's what the system should do. That's a function that we should have so we'll provide it.

Subject 2: Okay

Jia: Okay so there has been a face that's been recognized. You can go ahead and ask.

Subject 2: Who was recognized? (repeats 2 times)

Alvin: Who was recognized? (repeats)

System: "'Who was recognized?"' Carol.

Subject 2: Who was recognized?

System: "'Who was recognized?"' No one.

Subject 2: Who was recognized?

Alvin: You need to speak into the bluetooth.

System: "'Who was recognized?"' No one.

Andrew: Yeah, the recognition system is very dependent on what kind of lighting, so the ones we have for Hayato aren't the best

Jia: And he recently got a haircut, so...

Subject 2 $*$ to Jia $*$: Are you there?

Jia: No, I'm not in it.

Andrew $*$ to Lydia $*$: Move a little bit to the right

Subject 2: Who was recognized?

System: "'Who was recognized?"' Jia.

Lydia: Wrong person.

Subject 2: Huh.

Team FACE member: Okay, the face is found.

Subject 2: Who was recognized?

System: "'What was recognized?"' Jia.

Subject 2: How accurate have you found this to be?

Andrew: Okay, so if we take pictures today with these settings as people look right now, and they were to stand in the exact same spot where the pictures were taken, it would be perfect.

Subject 2: Yeah but they could have had on...

Andrew: Exactly, and that's kind of the trouble we've been running into

Subject 2: And can you solve this problem?

Andrew: Yeah, definitely. It's a solvable problem, it just requires more time on our part to refine our algorithms. There's tons of research in this area, we just have to pick the right method for this application.

Subject 2: I want one of these. To walk around the streets, you know? I really would volunteer to do this.

Member: Great!

Subject 2: I want one of these. I've been wanting this for years.

Hayato: When the system gets better, we'll definitely let you know

Subject 2: Fine, fine. You know, you ought to put in on an iPhone. Why don't you make an iPhone app?

Andrew: Well, okay, so one of the problems we faced with the iPhone is that it doesn't have enough computation power to get...eventually they'll get good enough to run it but, right now, we need a laptop. Yeah so it's just a technological limitation. But in a few years, that should be solvable.

Subject 2: Cool.

Kailin: (Subject 2), would you feel comfortable using this system in public?

Subject 2: Hell yeah!

Kailin: Okay, and in what setting do you see yourself using this in?

Subject 2: In my office.

Kailin: Okay.

Subject 2: If I'm looking for stuff and if somebody says, "'Hi, X,"' it's not like a sighted person where you can see so-and-so coming towards you and you've got like twenty feet to recognize that's so-and-so coming towards you. I don't get that. I just get, "'Hi, X."' And by the time the brain kicks in, I think, "'That's so-and-so,"' they're gone. And I'm thinking, "'I'm not sure, I think that's so-and-so."' But I don't want to call out to them in case it was somebody else. I'll make a fool of myself. So yeah, I want it at work.

Subject 2: The frequency might be a variable thing like do you want it feedback every 5 seconds or 30 seconds? 15 seconds?

Alvin: Ideally we'd want to give you feedback whenever the expression changes

Subject 2: Right, but if somebody's smiling continuously for twenty seconds, like when you're talking about something fun, you don't want it to be like, "'Smile. Smile. Smile. Smile."'

Alvin: If he's continously smiling, it'll only tell you, "'Begin smile."'

Subject 2: Well, it didn't before.

Jia: Both the beginning and the end, so when it changes back to neutral face it'll -

Subject 2: But it said, "'Smile. Smile. Smile,"' before.

Alvin: Yeah, that's a problem.

Jia: Well, we're trying to change it so that it only -

Subject 2: Or maybe it could say, "'Begin smile,"' and, "'End smile,"' "'Begin frown,"' and, "'End frown,"' or something and that way you would be getting feedback in between but you would know when it began and when it ended.

Hayato: Do you think hearing this feedback would change the way you would interact with, say, someone like me?

Subject 2: Hmm, that's a question...I don't think so. It might be nice to know, I mean, if someone's frowning, I might not come up to them and say, "'Hey, how are you?"' you know. "'Isn't it a great day?"' you know, and maybe if someone was smiling I wouldn't come up to them and say, "'Oh god, today's a crappy day."' You'd maybe want to tailor what you say and how you say it to the way they're looking. Maybe that would sort of tailor how I'd interact with them. But I don't know, I've never had this before, so I don't know.

Hayato: A question that I have is do you think in terms of like smiles or do you think in terms of whether somebody is just happy. Do you understand what I-

Subject 2: Yeah yeah. I don't really think about the physical facial expressions, so that doesn't really matter to me. In fact, whether somebody's happy or not would be more of what I'd be looking at.

Hayato: Ok.

Kailin: Alright do you have any more questions for (Subject 2)?

Hayato: No.

Kailin: Ok, do you, (Subject 2)? We have nine minutes left, so we want to start wrapping things up, but do you have anymore general feedback for us about the device?

Subject 2: I don't think so.

Kailin: Ok.

Subject 2: But I'd certainly be interested in continuing with the testing if you are looking for testers.

Kailin: Yeah definitely. Probably in the spring sometime when we have everything close to finalized.

Subject 2: You've got my contact information.

Kailin: Great. Thanks so much. So, we have some monetary compensation for you for coming all the way from where you came from.

Subject 2: You can just hand it to my cab driver.

Kailin: Hold on sorry, we're just getting it.

Hayato: So, thanks for coming. Here's 40 dollars.

Subject 2: Oh thank you very much.

Kailin: Thank you so much for...

Hayato: We promise this is 40 dollars.

Kailin: Thank you so much for taking your time.

Subject 2: What I can do...

Hayato: Yeah, (Subject 1) had one of those.

Subject 2: Yeah, I love this thing. Of course, I have one on my iPhone also, but...

Hayato: Oh.

Subject 2: If you're going to be blind, now is a pretty good time.

Hayato: So we'll definitely work on the facial recognition system.

Subject 2: I wonder how many faces you can put in. I mean, how much room on a hardrive does each face picture take up? I mean, like if you had a 20 gig

hardrive, you can have more than that, but if you did, I'm wondering how many faces.

Andrew: Oh, you can store tons of faces.

Alvin: Yeah, it depends on the algorithm you're using as well.

Hayato: Yeah. Today we really just wanted to get an understanding of potential users.

Subject 2: Actually there's a new thing on iOS 5 on the iPhone where you can take a picture and it'll tell you whether the face is centered on the viewfinder. So a blind person could take a picture and if they could get it from the iPhone into this the laptop or whatever, and that way you could take a picture of a person that you wanted to add to the database.

Andrew: Yeah. One thing that was actually suggested to us was to look at people's Facebook pages.

Subject 2: Oh.

Andrew: And use that. I don't know if-

Subject 2: See, I don't use Facebook.

Andrew: Ok. Yeah, so-

Subject 2: I don't do facebook.

Andrew: I think that's what some of the...Who told us that?

All: It was NFB. ∗ laughter ∗

Subject 2: That's fine.

Jia: But if the users that...If there's a person who uses Facebook who you interact with, it would still be possible to find their pictures on Facebook. So you wouldn't have to take that many pictures of them because there's essentially a ready made database of pictures.

Subject 2: Right, right.

Kailin: Ok, thanks so much (Subject 2).

11.2.3 Subject 3

Hayato: Hi. My name is Hayato, and I'm a member of Team FACE. Thanks again for participating in our experiments. Before we begin today, I just want to go over what our device consists of so that you know what you're using. There are 3 pieces to our device a Bluetooth to wear on your ear to obtain feedback (and give controls), a camera mounted on a cane, and a laptop to power the system. The camera allows you to identify people who are in front of you and the expressions they have on their faces. Do you have any questions so far?

Subject 3: Yeah, I just want to be clear. This system has a two-fold purposeyou can recognize people's faces, and it also tells you what their expressions are?

Hayato: Right. That's correct.

Doug: It compares the face to a set of images that we have stored locally on the computer here. And so you already have a picture of someone that you know and it will compare the picture on the camera to your set of pictures, and if it recognizes them, it'll tell you, "'Doug is here"' or, "'Hayato is here."'

Subject 3: Ok.

Rosie: Another piece of the device, in addition to the camera and the cane, is the Bluetooth. The Bluetooth will emit a sound when a person is in front of you and will convey certain facial expressions, like whether they're smiling or frowning. As for the experiments, we will be asking you to participate in four experiments today. In our first two experiments, you won't be using the

device, but you will be for the last two. We'll let you play around with the device during experiments three and four so you can become more familiar with it and its different features. Is that ok?

Subject 3: Ok.

Hayato: For the first experiment, we're going to play a video for you. We're going to ask you to identify when you think the speaker is smiling using a clicker. Rosie will explain how to use the clicker.

Subject 3: How will I know that? If they're smiling? It's just guessing right?

Rosie: You'll have to base it off of their voice and your intuition. So every time you think that the individual in the video is smiling, just press the clicker like this.

Rosie: Yea, exactly. That's perfect.

Subject 3: Will I be prompted to indicate when I hear a smile?

Carol: No. Whenever you think that the speaker is smiling, you just give it a click.

Rosie: There are no right or wrong answers.

Subject 3: So what if I don't hear a smile?

Rosie: Then you don't have to click. If you don't hear a smile during the entire video, you don't have to click at all.

Hayato: So just a little background on what the video is about. It's actually an interview. The interviewee is a renowned photographer who takes pictures of wild animals. So that's going to be the general topic of the conversation. And you will only be clicking when you believe that the interviewee is smiling. So, not the interviewer. Does that make sense?

Subject 3: Now, the photographer is the one conducting the interview?

Hayato: No.

Doug: He is the subject of the video.

Subject 3: Oh. He's the subject of the video. Just wanted to make sure

(Carol conducts sound/volume check)

Hayato: Ok. We're going to start the video now. Just pay attention and click away.

Subject: Ok.

(Video 1)

Hayato: Ok. That was our first video. For consistency purposes, we're going to play another video. It's of the same interview. And you will do the same as before - click when you think the speaker is smiling. Are you ready?

Subject 3: Sure. I'm ready.

(Video 2)

Hayato: So now you have finished the first experiment. The purpose of this experiment was to see how many smiles you could pick up without using any assistive devices.

Subject 3: Ok. I guess it's already done, but...so I was supposed to click each time I heard one? Because I did hear a couple, but I thought you were supposed to do it just once. Right?

(Note: subject 3 misunderstood the directions. Only thought he was supposed to click once total per clip even though he heard multiple instances of potential smiling.)

Rosie: Oh, why don't we redo it? We can repeat the experiments. So basically, click every time you think you hear a smile. So if I'm smiling right now, you click. And then I talk and talk, but if I smile again, you click again.

Subject 3: Oh! Ok.

Carol: So we'll restart from clip 1.

(Video 1)

Subject 3: The smile was not in the first one.

Rosie: Ok. We'll just do the second one.

Subject 3: Yea, the second one.

(Video 2 plays)

Hayato: Again, the purpose of this experiment was to see how many smiles you could pick up without using any assistive devices. So now, we're going to compare how you felt during this experiment to how you feel in this next experiment where you will be given feedback.

Doug: Can I see the clicker? Thank you.

Carol: For this experiment, you won't have to do any clicking. You just have to listen, and we'll ask you some questions concerning your feedback afterwards.

Subject 3: Ok.

Carol: This one is the one with chimes. So, you'll hear a chime whenever the person is smiling.

(Video with chimes)

Hayato: Ok. So now we're going to play another video. Now, the feedback is the word "'smile"' instead of a chime.

(Video 4 plays)

Hayato: So now that we've played the videos for you, with and without feedback, we have several post experimental questions.

Subject 3: Ok.

Hayato: So the first question is on a scale of 1-5 (5 being the highest), rate how obtrusive you found our feedback system.

Subject 3: Intrusive?

Doug: Was it distracting from the audio content? If you were to hear smile going on in your ear while you were talking to someone, would it be obtrusive to you? Like the whoop sound would that be annoying? Would it make it hard to listen to someone else while also having that whoop or smile going on.

Hayato: So if it was really distracting, you would give it a five. And if you didn't mind it at all, you would say one.

Subject 3: Well, it was kind of distracting. I'd say at least a two.

Hayato: Did the feedback give you any new information that you couldn't have otherwise conceived? Did you learn anything about the audio content from the feedback?

Subject 3: I'm not so sure it did, except that I disagreed with it. It didn't sound to me as if the person being interviewed was smiling. It sounded completely and totally serious. It did not sound as if [he was smiling]. Total seriousness is what I got.

Hayato: Ok. Did the feedback change the way you perceived the video?

Subject 3: No. It was puzzling because it didn't appear to me that the person was smiling.

Hayato: So again, on a scale of 1-5 (5 being the highest), rate how satisfied you are with our device or the system.

Subject 3: It is difficult to conclude something. Hmmm.

Carol: So if you had a device like this, would you be willing to use it? Or did you like how the feedback conveyed the expressions?

Subject 3: I don't think it's reliable.

Hayato: Would you like to expand on that?

Subject 3: I feel sort of uncomfortable talking about something that I'm just basing on my intuition and gut feeling. But back to the same thing that I said earlier, when the system was indicating "'smiles,"' it didn't seem that way to me and I disagreed with it. It's not something that I would consider reliable.

Hayato: With that said, how can we modify our device or system to better suit your needs?

Subject 3: Again, I'm in areas that I'm uncertain about, but the indicators could be changed, you know, like the beeps. In my mind, the beeps were more distracting than just the actual word "'smile."' And I had thought that one of the conductors here used the word "'chime"' and I thought, well ok, I have a pretty decent idea of what a chime sounds like. But when that "'bwoop"' or whatever it was

Doug: I suppose chime is a little bit of a misnomer for it. It was computer generated, and that's why it didn't necessarily sound like a wind chime.

Subject 3: I would possibly change it to maybe a "'ding"' or something like that.

Hayato: Yea, we have some of those. We just didn't implement them into the videos, but if you'd like to hear them and comment on them, that would be great.

Subject 3: Yeah, I was checking for something in my pocket to give you an

idea of what I like to hear. (subject 3 plays musical chime on one of his assistive devices) Something like that. Just one note. This is a pedometer. So when I walk and the music is on, it can become annoying. Just one chime or note, you know?

(Carol searches for the other chime options on computer)

Hayato (while waiting): Going back to what you said about our system and how it recognizes smiles, we were careful not to say, "'happy"' because smiles don't always translate to happiness. But what we're trying to do is to make a system that will tell you know when the person is actually smiling. And from there, it depends on how the user uses that information and incorporates it into in his or her interactions.

Subject 3: Based on what you just said, I thought, or in my mind, smiles mean happy. And that's what I was basing what I said on. Because the person being interviewed, it sounded like a straighter-head interview, and I couldn't detect from my own gut, happiness there, and that's why responded the way I did.

Hayato ∗ to Carol∗: Ready?

Carol: So these are some of the other sounds that we have.

(Carol plays chime options)

Subject 3: That's better.

Hayato: So we're also planning to have our system eventually be able to recognize an array of expressions, not emotions. So that includes when the person is frowning or making an angry face. So that's why you heard some lower pitched chimes. And those are potential indicators for negative expressions.

Subject 3: Ok.

Doug: So, would you say that some of the other chimes or tones would be

more helpful, or less intrusive?

Subject 3: Yeah. You should keep in mind that I'm just one person. All blind people don't think alike. Most of us hear an array of beeps and sounds during the day, whether we're at home or whether we're in our work atmosphere because in the technology that we utilize, there are often times beeps, bleeps, bells, whistles. So if there are some sounds that we hear all day, and we go to an experiment like this and that sound is played, it might tend to be distracting. I don't use that much technology, but I hear things when I turn on my computer. I hear beeps and bleeps when I'm on the internet. Even the rings in the elevator. When you hear it all day, it can become distracting. I don't know how you would determine this, but get a sound that you think that a person doesn't normally hear a lot.

(Carol's phone vibrates in the background)

Subject 3: That sounds like a phone is vibrating.

Carol: Yep. That's my phone. (Carol apologizes and steps out of the room)

Subject 3: Vibrating phones are necessary but they are annoying. I generally like bells or alarm sounds. I like clocks. I have too many of those. It's just something that I like. I don't know why. I've got my alarm set on both my cell phones. I have three alarm clocks and a watch. It's a gentle sound. It's not loud or threatening. Maybe that's unique with me. I don't know.

Hayato: Ok. So that concludes the first half of our experiments. The second half will involve our actual system so the cane and the camera itself. So before we get started on the last two experiments, we're going to give you some time to play around with it and familiarize yourself with it. Doug, you want to take over?

(Doug and Alvin help the subject 3 learn how to use the system)

Doug: So we have the cane here. It's actually pretty tall. You can feel it. And up at the top, here, this is the camera. And then you can feel behind it - there's a wire that goes to the computer. And we have the laptop here in front of you with speakers. That's what we're going to be playing the sounds off of.

Subject 3: Ok.

Doug: This is the Bluetooth headset. Are you familiar with Bluetooth headsets and how they work?

Subject 3: No, but I think they're wireless right?

Doug: Yea, they're wireless. And so this right here, that's an earpiece. Would you mind putting that in your ear? There we go. Is that ok? Is everything comfortable?

Subject 3: Yeah.

Doug: Alright. There are a few commands that we can use with the system. The first one is, "'Who is recognized?"' The camera isn't pointed to anyone right now, so that's why it just said, "'no one."' So if I turn and point the camera at me, and I say, "'Who is recognized?"' (Doug demonstrates, and system answers correctly) So it recognized me. Does that make sense? So the camera is pointed at me right now. (Doug repeats the command, and the system says Doug's name once again.) So now I'm going to point the camera at Hayato, and I'm going to say, "'Who was recognized?"' (Doug demonstrates, and system says "'no one"' at first, but on the second attempt says, "'Hunto"') Oh, that's just the way the system says Hayato.

Subject 3: Is this device in my ear supposed to be used in anyway?

Doug: That's actually the microphone. That's how the device hears what you're saying, it's through that Bluetooth headset. And it's actually picking up

my voice too, and that's why it's recognizing. Ok, I'm going to hand the cane back to you.

Hayato: So now, we're going to let you practice directing the device.

Doug: If you don't mind standing up, I can show you a little bit better.

Subject 3: Yea, that's probably more practical.

(Subject 3 learns/practices aiming the camera at Hayato with the help of Doug's instructions)

Doug: Does this make sense?

Subject 3: Yeah, I guess I would learn. More practice would make the pointing more accurate.

Doug: Oh yes, definitely. And I'm sure with five minutes everything won't be perfect. But we just want to get a general sense of where we can improve things with the device.

Doug $*$ to Hayato$*$: Do you want to the talk about the next experiments?

Hayato: So the third experiment is going to be kind of like a game. We're going to have three conductors - me, Harrison, and Doug. So we're going to stand here in a line. First, you're going to go down the line and point the camera at us it's going to tell you who we are. And then, we're going to shuffle ourselves, and you're going to have to use the device to figure out where each of us is standing. And then afterwards, we'll ask you a few questions.

Subject 3: Alright.

Hayato: Alvin's going to assist you with using the device.

Alvin: Ok. In the future, our system will tell you when a face is recognized. But right now, we don't have that feature, so I'm going to be that feature for you.

(First part of experiment 3 takes place)

Subject 3: Ok. So should I point to someone?

Alvin: Yea. Can you turn to the left a little bit? A little more

Subject 3: Really?

Doug: There you go. Now Hayato is in the frame.

Alvin: Can you bring it a little closer? You can now ask the system, Who is recognized?

Subject 3: Who is recognized?

(system correctly identifies Hayato)

Doug: There you go.

Subject 3: Hmmm

Doug: Now if you want to point it at me.

Alvin: Ok. To the left. Try now.

Subject 3: Who is recognized?

(System incorrectly recognized Doug as Hayato, even after multiple attempts)

Subject 3: It likes you.

Doug: Yea, it likes Hayato more than it likes me I guess. Ok, let's try Harrison.

(Subject 3 aims the camera at Harrison with Alvin's guidance)

Subject 3: Who was recognized?

(system correctly identifies Harrison)

Alvin: There we go!

Doug: I guess it just doesn't like me.

Hayato: Obviously, we're still working to improve the system.

Subject 3: Yeah, I like this. I like this.

Hayato: So now, we're going to get started with the "'game."' We're going to

shuffle ourselves and stop talking.

(Second part of experiment 3 takes place)

(Alvin provides assistance to help subject aim camera.)

(System correctly identifies Harrison.) (System incorrectly identifies Doug as Hayato) (System correctly identifies Hayato)

Hayato: So, we got two out of three.

Doug: Yeah, I was in the middle. Again, it didn't like me.

(Subject 3 laughs)

Doug: The other two were correct. Harrison was to your right. I was in the center, and Hayato (or "'Hunto"') was on your left.

Subject 3: So two out of three.

Doug: We're still improving things. Things improve from week to week. We have a few questions we want to ask you about this system.

Alvin: You can take the Bluetooth off now.

Doug: Yea, I can take the cane. You can sit down again.

Hayato: We're going to save the questions until after the fourth experiment.

(Laptop with script ran out of batteries. We had to delay questions until we got another laptop ready)

Hayato: So that was the face recognition part of the system. Now, we're going to play around a little bit more with the expression recognition system. This relates to the first two experiments that we did, but that was in more of a controlled setting. This time, we're going to use it in more of a natural setting. We're going to mount the camera onto a tripod and point it at my face. Meanwhile, we'll have a one-on-one conversation. We can talk about anything you want. Hopefully, the system will pick up on my expressions and, it will give

you feedback via the speakers. After this, we'll ask you the post-experimental questions.

(Alvin mounts camera and Hayato sits in front of subject 3)

Subject 3: I'm very good at talking about mundane things.

Hayato: Haha. Me too.

Subject 3: So is the system working now?

Hayato: No. He's still setting it up. Alright. Alvin just gave me the okay signal, so we can begin talking. Ok.

(System beeps)

Hayato: So that beep right there indicates a smile.

Hayato: How was the metro ride?

Subject 3: It was comfortable. Now I hear something in your voice, which I didn't hear in the voice in the [video] experiments.

Hayato: What did you hear?

Subject 3: Happiness.

Hayato: Do you have any hobbies, (Subject 3)?

Subject 3: Hmmm. Not really. Maybe I'll start collecting alarm clocks because I like the sound. So I'll just fill my house with alarm clocks and let them go off at the same time to amuse myself.

Hayato: Haha. Are these beeps distracting?

Subject 3: No, not really. Like I said, they could be for some people. I guess if I had an extremely gruesome day at work, I'd probably prefer to hear a bell. I thought it missed a smile there for a second.

Hayato: Sometimes, there's a lag. So, just like with the face recognition system, we're also working to improve this one. We're kind of running short on

time right now, so we're going to end the conversation there. Now I'm going to ask you the post-experiment questions.

Subject 3: Ok.

Hayato: These first questions will be about the face recognition system. For the following questions, I'm going to ask you to rate your opinion on a scale of 1-5 (5 being the highest). This recognition system was useful.

Subject 3: 4.

Hayato: This recognition system improved the ease of recognizing people.

Subject 3: Same number. 4.

Hayato: I would feel comfortable using this system in public.

Subject 3: 5.

Hayato: Ok. And just so you know, when we finalize this product, the laptop will be transported in a backpack. And it will be connected to the camera on the cane and you will receive feedback and give commands via the Bluetooth headset.

Subject 3: Ok.

Hayato: Next question. I would use this system in a group setting.

Subject 3: Very useful. 5.

Hayato: What was the most useful part of our device?

Subject 3: Focusing the camera.

Hayato: What was the least useful part of our device?

Subject 3: I'm not sure if there was a least. I don't have an answer for that.

Hayato: How can we modify our device to better suit your needs?

Subject 3: Umm, maybe the indicators.

Hayato: Just a reminder we're still talking about just the face recognition system.

Subject 3: Oh, ok. I'm not certain as to what kind of improvements...which is why...if you noticed I gave it very high ratings.

Hayato: You said you would use this device in a real life setting.

Subject 3: Yes, it would be very useful.

Hayato: Do you have any specific reasons?

Subject 3: Well I'm a social person. I like to mingle with people, and if I'm in a large group, I'd like to be drawn towards someone I know. I'm not going to know everybody in a crowded place, [but I'd like to find] people from my area or network.

Hayato: Do you have any other feedback for us concerning the face recognition system?

Subject 3: No, I don't think so.

Hayato: Ok. So now we're going to shift focus to the very last experiment, where you and I had our conversation. So this is about the expression recognition system. Once again, I'm going to ask you to rate on a scale of 1-5 (5 being the highest). Rate how obtrusive you found our feedback system.

Subject 3: 2.

Hayato: Does the feedback give you new information that you couldn't have otherwise known?

Subject 3: Perhaps, but it's hard to be specific.

Hayato: Again, I guess you had some trouble trusting the beeps. Would you say that was the case this time?

Subject 3: I'm getting used to it.

Hayato: Ok. On a scale of 1-5, rate the following statement: I would feel comfortable using this system in public. 5 meaning that you would be very

comfortable.

Subject 3: 5.

Hayato: Rate the following statement: This expression recognition system was useful.

Subject 3: I'll just comment before I give a rating. I'm not so sure about the facial expression recognition. I'd say, perhaps a 3.

Hayato: Rate the statement I would use this system in a group setting.

Subject 3: 3.

Hayato: Rate how satisfied you are with our device.

Subject 3: Overall, a 4.

Hayato: How can we modify our device to better suit your needs?

Subject 3: Maybe tweak the facial [expression] analysis.

Hayato: So it would be safe for me to say that, at this point, you would prefer the face recognition system over the expression analysis system?

Subject 3: Yes.

Hayato: If you don't mind me asking, what was the nature of your blindness?

Subject 3: Sure. I was blind at birth.

Hayato: I ask this because we're interested in seeing whether people who lost sight after birth have a preference for either system and vice versa.

Subject 3: Interesting.

Hayato: So do you have any other questions or comments for us?

Subject 3: Question. In order to recognize faces, don't you have to program data into the system about the person - how he looks or something? Like it was recognizing Doug or Harrison, did you put data in there about them or what? Did you put a copy of how they look in the machine or what? How does that

work?

Doug: I can explain that one. We do have pictures pre-compiled, already stored in memory on the computer. And in the future, what we're hoping to do is integrate our system with facebook so that you could pull somebody's public profile picture off of Facebook and use that to recognize them.

Subject 3: Ok. That's more practical. I was just wondering.

Doug: Ideally, you could either do that. Or say you have family members who don't have facebook you could use the system to take a picture of them before-hand and use that picture, in the future, to recognize them. If you went up to a stranger, you wouldn't be able to know who they are. So yes, to answer your question, there is already information about each of us stored in the system.

Subject 3: Ok. That's it.

Hayato: Ok. Thanks for coming again, (Subject 3).

11.2.4 Subject 4

Note: The computer abruptly stopped recording halfway through the experiment. It stopped around the 25th minute, while the experiment lasted approximately 1 hour.

Hayato: Hello. My name is Hayato, and I'm a member of Team FACE. Thanks again for participating in our experiments. Before we begin today, I just want to go over what our device consists of so that you know what you're using. There are 3 pieces to our device a Bluetooth to wear on your ear to obtain feedback (and give controls), a camera mounted on a cane, and a laptop to power the system. The camera allows you to identify people who are in front of you and the expressions they have on their faces. The Bluetooth will emit a sound when a

person is in front of you and will convey certain expressions, like whether they're smiling or frowning. We will be asking you to participate in four experiments today. In our first two experiments, you won't be using the device, but you will be for the last two. We'll let you play around with the device during experiments three and four so you can become more familiar with it and its different features. For the first experiment, we will be playing a video for you. We will ask you to identify when you think the speaker in the video is smiling using a clicker.

(Alvin hands the subject the clicker and allows her to practice clicking it.)

(Hayato gives background information about the topic discussed in the videos.)

Hayato: So every time you think the person being interviewed in the video is smiling, just press the clicker. Disregard the expressions of the interviewer in the video.

(Sound/volume check)

(Subject 4 is ready.)

(Video 1)

Hayato: Ok. So that was the first video. For consistency purposes, we will play a second video (but a part of the same interview), and we will ask you to do the same as before click when you think the interviewee is smiling.

(Video 2)

Hayato: Now, you have finished the first experiment. The purpose of this experiment was to see how many smiles you could pick up without using any assistive devices, and then to compare how you felt during this experiment to how you will feel in the next experiment where we'll you'll be given feedback via speakers.

Subject 4: Ok.

Hayato: So this brings us to the second experiment. We'll be playing two more videos (same interview), but this time you'll receive feedback as to when the interviewee is actually smiling. In the first video of this second experiment, you will hear a bleep-like sound when the person is smiling. You just have to sit back and listen.

(Video with chimes)

Hayato: So that was the third video. For the fourth and final video, instead of hearing a bleep, you will hear the word smile in the background whenever the person is smiling. After this video, we'll ask you a series of post-experimental questions.

Subject 4: Ok.

(Video with "'Smile"')

Hayato: Ok. So now, I'm going to ask you a series of post-experimental questions. On a scale of 1 to 5 (5 being the highest), rate how obtrusive you found our feedback system.

Subject 4: With the smile or the beep?

Hayato: Both.

Subject 4: I found it fascinating, in one way, that the guy was smiling so much. But as far as being obtrusive, I'd have to say...you can't quite pay attention, so I'd say maybe 4.

Hayato: Ok. And just to be clear, the system tells you when the person is actually smiling, but smiling doesn't always translate into happiness. So the

second question: Does the feedback system give you new information that you could not have otherwise conceived?

Subject 4: This is a yes or no question right?

Hayato: Yeah.

Subject 4: I'd say yes.

Hayato: And if so, what new information did you gain and did it change the way you perceived the video?

Subject 4: Having just listened to the voice without knowing the smile part, I wouldn't have guessed that he was smiling so much. So it was helpful.

Hayato: Can you explain how it was helpful to you?

Subject 4: It clued me into what his facial expression was, which I wouldn't have gotten otherwise.

Hayato: Ok. On a scale of 1 to 5 (5 being the highest), rate how satisfied you are with this system.

Subject 4: I'd say a 3. It balances out in that it's helpful but it's still disturbing.

Hayato: How can we modify this system to better suit your needs?

Subject 4: I don't know. And I don't know whether it's better to have the "'smile"' or the beep. Maybe less of a "'whoop"' and more of like...something like an email in your box.

Doug: Like a ding?

Subject 4: Yeah. I think that "'whoop"'...it kind of cuts out part of the actual conversation, so I was missing some of it. And I actually wasn't paying attention to the content.

Hayato: Ok. So that concludes the first half of the experiments. In the

second half, the experiments will involve the actual system itself. The first two experiments were meant to mimic what our actual system does.

(Doug begins to set up the system and explains in detail its hardware components.)

Doug: Are you familiar with Bluetooth headsets at all?

Subject 4: No.

Doug: Bluetooth is wireless technology, and so there's a microphone and speaker.

(Doug shows the Bluetooth headset to the subject and explains how it works.)

Subject 4: I wear a hearing aid, so I'll take this out. It'll be fine.

Hayato: In these experiments, wearing the Bluetooth won't be necessary. We just wanted to show you that it would be included in our final system.

Subject 4: Let me try it this anyway. I'd be curious to hear what it sounds like.

(Doug continues to set up face recognition system.)

Subject 4: Did you folks do research on what types of canes to use? Or did you just pick this one particular cane?

Hayato: We didn't do too much research about it. What happened was that we had a meeting with the National Federation of the Blind in Baltimore, and they gave us this cane to work with. And we went along with it because it's pretty tall, and it's level with most people's faces. But we've come to realize that it's not the most widely used [type] of cane out there.

Subject 4: Right.

(While still waiting for system to start up, Hayato talks to the subject about Dr. Tang's approach to mounting the camera on backpack straps. Hayato also

discusses the major pros (convenience) and cons (difficulty in leveling camera with another person's face) of such an approach.)

(Doug explains the commands for using the face recognition system.)

(Doug assists subject in playing around with the face recognition system. Significant technical difficulties occur. System does not recognize Hayato, Doug, or Harrison.)

AUDIO RECORDING ABRUPTLY STOPS AFTER APPROXIMATELY 25 MINUTES.

11.2.5 Subject 5

Carol: So for our records, this is (Subject 5) and today is November 18, 2011. And so thank you again (Subject 5) for participating in our experiments. Before we begin today, I just want to go over what our device consists of so that you know what you're using. There are three pieces to our device: a bluetooth for you to wear on your ear to obtain feedback and for you to give commands to the computer, so it'll know what to do, a camera mounted to a cane, and a laptop in a backpack to power the system. The camera allows you to identify people who are in front of you and the expressions that they have on their faces. The blue tooth will emit a sound when a person is in front of you and will convey certain expressions like when they're smiling or if they're frowning. Right now our device can do smiling and surprise. Alright so we'll be asking you to participate in four experiments today. In the first two experiments you won't be using the device but you will for the last two And we'll allow you to play around with the device before the experiments, the last two experiments so you'll be more familiar with it and the different features that our device has. So for the first experiment,we'll

be playing a video for you and we'll ask you to identify when you think the person is smiling, using a clicker and Doug will give you the clicker.

Doug: Which hand would you prefer, right or left?

Subject 5: Left.

Doug: Ok, this is the clicker and then underneath your thumb is a button and you can press it right now to see how it works. So every time you hear a smile, you'll want to click, or think that the person, uh, subject in the video is smiling.

Subject 5: Ok.

Doug: I'm just going to reset it really quick.

Carol: So everytime, like Doug said, everytime you think the individual is smiling. And so for consistency purposes, we'll be playing two videos for you and you'll be doing the same thing.

Carol: So this is the first video, so do you have any questions before we start? So click every single time you think that the person is smiling.

Subject 5: Ok.

Doug: Just so you know beforehand, this is an interviewer with an interviewee and we're more concerned with the interviewee, so the person answering the questions.

Subject 5: Ok.

(Video 1)

Subject 5: The person sounds so serious.

(Continue Video 1)

Carol: Alright so that was the first video and then we'll be playing the second video.

Doug: I'm going to start the second video now.

Carol: So when you think the interviewer is smiling, click the clicker.

Subject 5: Just from the tone of voice, it just sounds like the person is very serious. And I just figured the person would have had a serious expression on their face. I don't know if I got it right. Did I do it right, or not?

Doug: Oh are we allowed to...oh not yet, but we can tell you after the second video.

Subject 5: Oh.

Doug: Ok, I'm starting the second video now.

(Video 2)

Doug: Can I see the clicker? Just to record, there was one click.

Carol: So that was the end of the first experiment and the purpose of that experiment was to see how many smiles you picked up without using any assistive devices. And later on, we'll ask you to compare how you felt during this experiment, to how you will feel in the next experiment we will be giving you the sounds for when the person is smiling. So this is the second experiment and for experiment two, we'll be playing two more videos but this time you'll be able to hear when....you'll know when the person is smiling. So we're going to show you, or play samples of what to expect when you're watching the videos.

Subject 5: Ok.

Carol: So um, you'll hear this sound when you hear a smile.

Doug: I don't have that sound.

Carol: Oh, I don't think it, wait is it the same sound?

Doug: Oh, it's actually still on your flash drive.

Esmaeel: Here we go I have it

Carol: So basically this will be the sound you hear whenever the person is smiling.

Doug: Uhhh.

Carol: You can just play from a different video....

Doug: Oh ok. So these are the chimes and they're more of a ...dont think of windchime.

Carol: So those "'doo-doots"' are what you're going to hear.

Subject 5: Oh! That doesn't sound like a chime!

Carol: Haha yeah, we used the wrong word!

Doug: It's more of a....it's like a computer-generated sound

Subject 5: Wow, I sure blew that one!

Carol: But yeah, so that would be the sound you'd hear. But for the second video, for the sound, you'll hear someone saying smile.

Doug: Yea so here's the second clip that you heard before with the smile added.

Subject 5: Ok.

Carol: So you'll hear "'smile"' every time the person is smiling.

(Play "'Smile"' noise)

Doug: Ok so that's an example.

Carol: So then, during these videos when you're watching them, you don't have to do anything, we're just going to ask you a series of post experiment questions afterwards to get your feedback on the system. Do you have any questions?

Subject 5: Ok.

Carol: So you don't have to do anything for these videos, just listen.

(Video with chimes)

Doug: Ok, now we're going to play another video with the smile noise.

Subject 5: Ok.

(Video with "'Smile"')

Doug: Ok.

Carol: So now we have post experiment questions for you. So on a scale of 1-5, 5 being the highest, rate how obtrusive you found our device. So, the feedback system would be the smile, the person saying smile or the sound.

Subject 5: I think both of them were pretty effective, I think.

Carol: How obtrusive did you find it, so was it annoying or was it...

Subject 5: No, I didn't find it annoying. You mean as far as, ask the question again. I'm sorry

Carol: On a scale of 1-5, 5 being the highest, rate how obtrusive you found our feedback

Subject 5: Oh obtrusive, obtrusive. Um..,

Carol: So 5 being very obtrusive and 1 being not very.

Subject 5: No not.

Carol: So, 1?

Subject 5: Yeah.

Carol: Does the feedback give you any new information that you could not have otherwise conceived?

Subject 5: Yes.

Carol: If so, what new information did you gain and did it change the way you perceived the video?

Subject 5: Well, I think it did in the sense that when I did the first one I only clicked it one time and this gave me more information as far as what the facial expression was.

Carol: And on a scale of 1-5, 5 being the highest, rate how satisfied you are with the feedback system.

Subject 5: I'd give it 5.

Carol: And how could we modify our device to better suit your needs.

Subject 5: Basically different type sounds for different facial expressions would be good.

Carol: Now we're going to move on to experiment 3. You're going to have some time to play with our device-the cane.

Subject 5: Okay.

Carol: And to familiarize yourself with it.

Doug: Esmaeel, are we ready over there?

Esmaeel: I think we're going to have to go with just the expressions.

Doug: Just expressions today? Okay.

Esmaeel: Yea we're having from trouble with the face recognition.

Carol: So, we're not going to do experiment three?

Doug: Yea, we can skip experiment three. Okay, but she still needs to be familiarized the with the device, the features.

Subject 5: I mean I can come back, that's not a problem.

Carol: Cause I think our...

Esmaeel: Harrison can just show how it works.

Carol: If you could stand up.

Subject 5: Okay.

Carol: Here is the cane, and feel the button right here. That is how you know the camera is facing forwards.

Melissa: So the camera is mounted to the top of the cane.

Subject 5: Okay.

Carol: Alvin will tell you how to use the cane.

Alvin: So do you know which way it's pointing?

Doug: There's a dot there.

Subject 5: The button is straight ahead, so...Oh, ok it was like at a diagonal. Sorry about that.

Carol: It's fine. Sorry. So what are the commands?

Esmaeel: So you can ask, "'Who was last recognized?"' and it will tell you who the last person seen was. And then there's options for changing from chimes to words. What else do we have? That's basically it and the expressions are output automatically. So, its loading.

Carol: You can sit back down while we're waiting.

Subject 5: Ok.

Esmaeel: Sorry, it works with Harrison and that's all we need. Ok let's turn on the sound.

Alvin: We're ready if you wanna stand up again.

Subject 5: Okay.

Alvin: So you can ask the computer, "'Who was recognized?"'

Subject 5: What am I looking at?

Esmaeel: You can just ask, "'Computer, who was recognized?"'

Subject 5: Computer, who was recognized?

Esmaeel: (louder) Computer, who was recognized?

Computer: "'Computer, who was recognized? Harrison."'

Esmaeel: So...Computer, who was recognized?

Computer: "'Computer who was recognized? Harrison."'

Esmaeel: So it says Harrison's name. And if you try it on for example Alvin, it actually won't recognize Alvin.

Subject 5: Computer, who was recognized? Computer, who was recognized?

Esmaeel: Computer! Can you try it on me? Turn it towards me. A little further towards me. Computer, who was recognized? Computer, who was recognized?

Computer: "'Computer, who was recognized?"' No one.

Carol: So it doesn't recognize anyone other than Harrison at this point.

Subject 5: Ok.

Esmaeel: But it recognizes that I'm not Harrison and that's what we're trying to show right now.

Carol: So if you point it at Alvin who is right in front of you right now.

Esmaeel: So if Alvin smiles...

Computer: Smile.

Esmaeel: And we can try it again

Computer: Smile.

Esmaeel: And the smile works on multiple people, so I can stand in front.

Subject 5: So is the button here just to line it up?

Esmaeel: Right, just to guide you.

Subject 5: Ok.

Computer: Smile, smile.

Esmaeel: There we go. If we enable the other expressions as well, we could get surprise if I raise my eyebrows, or if I open my mouth it will say I'm surprised and it will do the same thing for virtually anybody else. And that's how it works!

Carol: So I think because our recognition isn't working today, we're going to skip experiment three where you would be recognizing a person. So we're going to carry on to experiment four.

Doug: In experiment four we use the tripod.

Michelle: Oh, do we?

Doug: Yeah.

Carol: And so we're going to be using the tripod instead of the cane.

Doug: Things are always a work in progress.

Esmaeel: The cable came out of the camera.

Doug: We should really invest in the screw top one.

Carol: So we're setting up the camera right now.

Subject 5: I'm going to get my phone out now, in case metro access does call.

Carol: It's 2:35 right now.

Michelle: So how long does this experiment take?

Doug: 5-10 minutes.

Michelle: 5-10 minutes, ok.

Doug: Can you turn it up by 1 or 2 notches Esmaeel-the volume.

Esmaeel: Yeah.

Michelle: Alright so for experiment 4, this is the experiment where we're going to use some of the expression analysis and to make you more comfortable with the device in a natural setting, we're going to let you use the device in a

one-on-one face conversation.

Subject 5: Ok.

Michelle: We'll have a one-on-one conservation on various topics. We'd like you to analyze our system while we're speaking and give us any feedback about our device. We'll be using the chimes for this experiment instead of the words. So do we need to change that? Would you prefer the chimes of the words?

Computer: "'Computer change feedback."'

Subject 5: It doesn't matter. Either one.

Alvin: Change to chimes. Computer change to chimes. Change feedback to chimes.

Computer: "'Change mode to chimes."' Mode changed to chimes.

Michelle: Now when I smile...

Computer: ∗ Smile sound ∗

Michelle: It will use a chime. So now the camera is pointed at me so it is going to evaluate my expressions as we're talking.

Subject 5: Ok.

Michelle: Okay. Ok so how about you just tell me a little bit about yourself.

Subject 5: Well I like to sing, I have three children who are grown and five grandchildren 6 and under.

Michelle: That's a lot of kids...hahaha. Um so what is your favorite childhood memory?

Subject 5: My favorite childhood memory...Going to the park with my granny and going on the swings and the playground. They had these swings like animals, wooden animals that you could sit on and that was fun.

Michelle: What do you like to do for fun, I know you said you like to sing....but

anything else?

Subject 5: Yeah. I like to sing, I like to read, I like to listen to music a lot, go out to eat with friends.

Michelle: Um so you can ask me any questions as well. Like anything you wanna know about me?

Subject 5: Well.

Michelle: Or any of us here?

Subject 5: What is your major?

Michelle: I'm a mechanical engineering major.

Subject 5: So why did you decide to come up with a device like this? And an experiment like this?

Michelle: Well we really wanted to come up with something that would benefit people in society and we saw a definite need in improving I guess communication and helping out blind people. We'd really like to get your feedback and integrate you guys into the process as well so we make sure we make a device that's exactly what you guys wants as well.

Subject 5: Bravo!! I give you an A+, everybody here. So with your major, is it a 4 year degree?

Michelle: Yup, 4 years and we're all seniors so one last semester.

Subject 5: I'm trying to think of what else to ask...haha. So what are your ambitions and goals for after you graduate and get a job

Michelle: Um.

Subject 5: What do you see ahead for you?

Michelle: Well I'm planning to go to graduate school and get my PhD in microrobotics for medical applications and then I see myself as a research professor

at a university like this. So I basically want to stay in school for the rest of my life...hahah. So now you said you had how many children?

Subject 5: I have three, I have Justin's my oldest, he was born in '80, Lauren was born in '82, and Michael was born in '83. And they're all in the computer field and my grandbabies are all 6 and under. The oldest, she's in first grade, and then Sophia, she's gonna be 4 in December, and she's learning to draw and write her letters and Samuel, he's two, gonna be 2 in December. and Michael and Lizzie have two. And their oldest's name is Caleb and he's two, and the youngest is 8 months old and his name is Cole. And I'm going to be there this weekend. So I'll get to watch babies and get on the floor and play and take them treats. It's more fun to be a grandma because you can relax and just enjoy and when its time to go you, go, go home.

Michelle: Alright thank you and now I'm going to ask you a couple of post experiment questions.

Subject 5: Ok.

Michelle: Alright! Um so for the following questions we're going to ask you to rate your option on a scale of 1 to 5, whith 5 being the highest, and rate how obtrusive the feedback system is, with 5 being the highest.

Subject 5: I think it's efficient as far as telling me when somebody is smiling.

Michelle: So how obtrusive did you find the chimes? Very obtrusive, not obtrusive at all...or somewhere in the middle?

Subject 5: Oh so, how obtrusive, that was the question?

Michelle: Yes.

Subject 5: Oh I was going to opposite way in my head. I wanna make sure I get it right. Yeah, a 1.

Michelle: Does the feedback give you new information that you could not have otherwise known?

Subject 5: I'm gonna say yes. I mean sometimes you can tell when someone is smiling from their tone of voice. But sometimes you can miss it.

Michelle: If so, what new information did you gain?

Subject 5: I believe that it showed me that the person was smiling more than I may have realized. I think its good to have the feedback because it gives you more insight into facial expressions.

Michelle: Alright, on a scale from 1-5, 5 being the highest, please rate the following comments. I would feel comfortable using this system in public.

Subject 5: Yes I would.

Michelle: So on a scale of 1-5, 5 meaning you take it with you every single time you go out, 1 being you always leave it at home.

Subject 5: I'd give it a 5. I'm wondering though is you know because everything is small, would it be possible to come up with a hand held device that you could have a small camera attached to?

Doug: I think in the future, potentially yes, but as of now..so. The computing power isn't enough to do it in real time.

Subject 5: Gotcha.

Esmaeel: Computers are moving in that direction though, so soon enough. Smart phones now are moving towards computers so in the next 5 to 10 years...we're getting there.

Michelle: So as of now, the system is the camera on top of the cane that you were holding earlier and then that links to a laptop that you would conceivably carry in your backpack, would be the whole system as it stands today.

Subject 5: Gotcha. Would it be able to recognize objects other than people's faces.

Doug: That's something. We're actually the continuation of a team. There was an undergraduate team before us and that was their main focus, that and also navigation. So their system was actually able to identify dollar bills versus twenty dollar bills and if you marked like a pencil, with a bot or striped pattern, you could recognize that pattern against the background, so if you drop your pencil, it would be able to recognize your pencil on the ground and direct you to it. But that's actually not the focus of our...

Subject 5: Oh, I know its not, I was just sort of asking you know in the future if it would be possible to combine all those things.

Doug: Oh yes, it's definitely possible.

Esmaeel: Actually one of our mentors, that's his goal. So he's trying to get all of these research projects like this he wants to package them and combine object recognition, navigation system and the social aides that are facial recognition and expression recognition all into one package.

Subject 5: That would be awesome! And that would be neat too if you could use it when you would walk with the cane and it would still take the pictures.

Esmaeel: And recognize things?

Subject 5: Yeah.

Michelle: Alright so on a scale of 1-5, how comfortable would you be using this system in public?

Subject 5: I would use it, I wouldn't find any problem with it, I'd give it a 5.

Michelle: Alright. Again, on a scale from 1-5: the recognition system was useful? I mean the recognition system that at least could tell Harrison versus

not-Harrison for now. Eventually on most days it can actually tell people apart. So how useful would you find that on a scale 1-5.

Subject 5: On a scale of 1-5, I would say a 5. That would be really important.

Michelle: Alright, so the expression system improved the ease of recognizing people.

Subject 5: Say that again.

Doug: Oh, Michelle, we didn't do this one, we just did the expressions.

Michelle: Yeah, sorry.

Doug: That was related to the experiment that we skipped.

Subject 5: But yeah I can still see the relevance of it and how important it would be to have recognized different people and stuff.

Michelle: Ok again on a scale of 1-5, I would use this system in a group setting.

Subject 5: Yes, 5.

Michelle: Alright, again on a scale of 1-5, rate how satisfied you are with our device.

Subject 5: Well because of the thing not being able to recognize who the people were, you know as far as the separation goes, I would give it a 4.

Michelle: What was the most useful part of the expression system?

Subject 5: I'd say being able to tell the facial expressions - whether they were smiling. I thought that was important.

Michelle: What was the least useful part of the expression system?

Subject 5: (pause)

Michelle: Was the anything you didn't like about it?

Subject 5: No I can't think of anything. The chimes or the words, I could

have used either one.

Doug: Did you like the chimes in the expression system versus that "'bloop"' sound?

Subject 5: I liked the chime better.

Doug: Ok that's what a lot of people have been saying?

Michelle: How could we modify our device to better suit your needs?

Subject 5: Just for this particular experiment, you'd have to make sure that, you'd have to have some sort of computer set up to be able to operate it out right? Would it have to be hooked up to a laptop or some thing? So that you could use it if you were going somewhere?

Michelle: Yeah.

Subject 5: I'm asking that you know if you were navigating with it or if you were at a particular location you'd have to be hooked up to a laptop and would you have to take a laptop with you in order to be able to use that.

Michelle: For right now, yes you would have to carry the laptop, but you wouldn't have to have it open or anything, it could run closed in a backpack so you kind of set it and forget it kind of thing. Just put it in your bag and you could carry it around. But in the future hopefully, it would be on a much smaller device like a cell phone or something.

Subject 5: Yeah, I hear you. I mean, I would use it, but you know if I had it that would be a lot of things to carry like all together. The computer, the cane, the folding cane, but I guess you wouldn't use it until you got to your final destination so if the cane collapsed, or does it collapse at all?

Michelle: The one we have right now doesn't but um the way the camera is actually mounted on top of the cane, the mount can be put on any cane so you

could put it on your foldable cane and use it that way.

Subject 5: Oh, alright!

Michelle: But for right now it can't do that, but in the future.

Subject 5: I would say have some kind of a cane that collapsed so you could put everything in the backpack and carry it to the destination that you're going.

Michelle: Also the camera right now can also mount itself on a standard tripod like you can go to the store and just buy a tripod for a digital camera and it can also be mounted on there. So if you were like in a meeting setting you could just take it out and put it on tripod and have it there instead of putting it on a big five-foot cane.

Subject 5: Gotcha.

Michelle: Alright so do you have any further comments of questions?

Subject 5: I can't think of any. I really enjoyed myself.

Michelle: Alright, thank you.

BIBLIOGRAPHY

[1] M. Argyle, V. Salter, H. Nicholson, M. Williams, and P. Burgess, "The communication of inferior and superior attitudes by verbal and non-verbal signals," *British journal of social and clinical psychology*, no. 9, pp. 222–231, 1970.

[2] "Visual impairment and blindness," *World Health Organization*, 2009.

[3] S. Krishna, G. Little, J. Black, and S. Panchanathan, "A wearable face recognition system for individuals with visual impairments," pp. 106–113, 2005.

[4] A. F. for the Blind, "Statistical snapshots," 2011.

[5] D. Matsumoto and B. Willingham, "Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals," 2009.

[6] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.

[7] E. F. Ersi and J. S. Zelek, "Local graph matching for face recognition," *IEEE Workshop on Applications of Computer Vision*, p. 3, 2007.

[8] K. Mikolajczyk, R. Choudhury, and C. Schmid, "Face detection in a video sequence - a temporal approach," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, pp. 96–101.

[9] S. Caperna, C. Cheng, J. Cho, V. Fan, A. Luthra, B. O'Leary, J. Sheng, A. Sun, L. Stearns, R. Tessler, P. Wong, and J. Yeh, "A navigation and object location device for the blind," 2009, gemstone Team Vision thesis.

[10] T. McDaniel, S. Krishna, V. Balasubramanian, D. Colbry, and S. Panchanathan, "Using a haptic belt to convey non-verbal communication cues: During social interactions to individuals who are blind," pp. 13–18, 2008.

[11] R. Yang and Z. Zhang, "Model-based head pose tracking with stereovision," 2002, stereo vision.

[12] S. Treuillet and E. Royer, "Outdoor/indoor vision-based localization for blind pedestrian navigation assistance," *International Journal of Image and Graphics*, vol. 4, pp. 481–496, 2010.

[13] L. A. Johnson and C. M. Higgins, "A navigation aid for the blind using tactile-visual sensory substitution." in *Proceedings of the 28th IEEE EMBS Annual International Conference*, 2006.

[14] A. J. O Toole, "Face recognition algorithms surpass humans matching faces over changes in illumination," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 9, p. 1642, 2007.

[15] R. Beveridge, B. Draper, G. Givens, and W. Fisher, "Introduction to the statistical evaluation of face recognition algorithms," 2005.

[16] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *Acm Computing Surveys*, vol. 35, no. 4, pp. 399–459, 2003.

[17] Q. Xiao and X. D. Yang, "A facial presence monitoring system for information security," *IEEE Workshop on Computational Intelligence in Biometrics: Theory, Algorithms, and Applications*, pp. 69–76, 2009.

[18] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[19] M. Nusseck, D. W. Cunningham, C. Wallraven, and H. H. Bulthoff, "The contribution of different facial regions to the recognition of conversational expressions," *Journal of Vision*, vol. 8, no. 8, 2008, nusseck, Manfred Cunningham, Douglas W. Wallraven, Christian Buelthoff, Heinrich H.

[20] T. Kanade, Y. Tian, and J. Cohn, "Comprehensive database for facial expression analysis," 2000.

[21] Y. Yacoob and M. Black, "Recognizing facial expressions in image sequences," 1997.

[22] S. u. Rehman and L. Liu, "Vibrotactile rendering of human emotions on the manifold of facial expressions," *Journal of Multimedia*, vol. 3, no. 3, pp. 18–25, 2008.

[23] S. Krishna, D. Colbry, J. Black, V. Balasubramanian, and S. Panchanathan, "A systematic requirements analysis and development of an assistive device to enhance the social interaction of people who are blind or visually impaired," 2008.

[24] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991, pp. 586–591.

[25] T. Ojala, M. Pietikainen, and D. Harwood, *Performance evaluation of texture measures with classification based on kullback discrimination of distributions*. IEEE, 1994, vol. 1.

[26] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[27] G. Agam, "Introduction to programming with opencv," 2006.

[28] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J. J. Gross, M. E. Jabon, C. A. C. Hutcherson, C. Nass, and O. John, "Real-time classification of evoked emotions using facial feature tracking and physiological responses," *International Journal of Human – Computer Studies*, vol. 66, no. 5, pp. 303–317, 2008.

[29] B. J. Balas, "Region-based representations for face recognition," *ACM transactions on applied perception*, vol. 3, no. 4, p. 375, 2006.

[30] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, vol. 1, no. 6, pp. 22–35, 2006.

[31] F. Bourel, C. Chibelushi, and A. Low, "Robust facial feature tracking," pp. 232–241, 2000.

[32] R. Brunelli and P. Tomaso, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042 – 1052, 1993.

[33] J. Bruno, "Face recognition," p. 3, 2010.

[34] J. F. Bugental and E. K. Bugental, "A fate worse than death: The fear of changing," *Psychotherapy: Theory, Research, Practice, Training*, vol. 21, no. 4, pp. 543–549, 1984.

[35] J. F. T. Bugental and E. K. Bugental, *Resistance to and fear of change*, ser. The Hatherleigh guides series. New York, NY US: Hatherleigh Press, 1996, pp. 33–46.

[36] R. Burns, *Structured interview and questionaire surveys*, 2000, ch. 30, pp. 566–593.

[37] R. Chellappa, "Lecture on face recognition," 2009.

[38] J. W. Dally, *A Product Development Process*, 4th ed. College House Enterprises, LLC, 2000, ch. 12, pp. 233–254, product Design Reference.

[39] N. Edwards, J. Rosenthal, D. Moberly, J. Lindsay, K. Blair, S. Krishna, T. McDaniel, and S. Panchanathan, "A pragmatic approach to the design and implementation of a vibrotactile belt and its applications," 2009.

[40] B. Froba and C. Kublbeck, "Face tracking by means of continuous detection," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, . C. o. C. V. Workshop and P. Recognition, Eds., 2004, pp. 65–72.

[41] J. Gothe, S. Brandt, K. Irlbacher, S. Rricht, B. Sabel, and B. Meyer, "Changes in visual cortex excitability in blind subjects as demonstrated by transcranial magnetic stimulation," 2002.

[42] A. Graziano and M. Raulin, *Research methods: a process of inquiry*, 7th ed. Boston, MA: Allyn and Bacon, 2009.

[43] C. K. Hsieh, S. H. Lai, and Y. C. Chen, "An optical flow-based approach to robust face recognition under expression variations," *Ieee Transactions on Image Processing*, vol. 19, no. 1, pp. 233–240, 2010.

[44] C.-L. Huang and C.-W. Chen, "Human facial feature extraction for face interpretation and recognition," *Pattern Recognition*, vol. 2, pp. 204–207, 1992.

[45] G. B. Huang, M. Narayana, and E. Learned-Miller, "Towards unconstrained face recognition," *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pp. 1–8, 2008.

[46] S.-H. Jeng, H.-Y. M. Liao, C.-C. Han, M.-Y. Chern, and Y.-T. Liu, "Facial feature detection using gemotetrical face model: An efficient approach," *Pattern Recognition*, vol. 31, no. 3, pp. 273–282, 1998.

[47] J. S. Ju, E. Ko, and E. Y. Kim, "Eyecane: Navigating with camera embedded white cane for visually impaired person," 2009.

[48] S. Kaiser, T. Wehrle, and S. Schmidt, "Emotional episodes, facial expressions, and reported feelings in human-computer interactions," pp. 82–86, 1998.

[49] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," pp. 506–513, 2004.

[50] N. Kela, A. Rattani, and P. Gupta, "Illumination invariant elastic bunch graph matching for efficient face recognition," p. 42, 2006.

[51] R. A. Krueger and M. A. Casey, *Focus Groups: A Practical Guide for Applied Research*, 4th ed., 2009.

[52] Y. H. Kwon and N. D. Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999.

[53] M. Lee and C. H. Park, "An efficient image normalization method for face recognition under varying illuminations," pp. 128–133, 2008.

[54] S. W. Lee and S. H. Moon, "Face recognition under arbitrary illumination using illuminated exemplars," vol. 40. Pergamon-Elsevier Science Ltd, 2007, pp. 1605–1620.

[55] B. X. Li and R. Chellappa, "Face verification through tracking facial features," *Journal of the Optical Society of America a-Optics Image Science and Vision*, vol. 18, no. 12, pp. 2969–2981, 2001.

[56] P. Li, H. Ai, Y. Li, and C. Huang, "Video parsing based on head tracking and face recognition," pp. 57–64, 2007.

[57] J. J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li, "Automated facial expression recognition based on facs action units," in *Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 390–395.

[58] J.-J. J. Lien, T. Kanade, J. Cohn, and C.-C. Li, "Detection, tracking, and classification of action units in facial expression," *Robotics and Autonomous Systems*, vol. 31, no. 3, pp. 131–146, 2000.

[59] D. G. Lowe, "Object recognition from local scale-invariant features," pp. 1150–1157, 1999.

[60] Y. M. Lui, D. Bolme, B. A. Draper, J. R. Beveridge, G. Givens, and P. J. Phillips, "A meta-analysis of face recognition covariates," pp. 1–9, 2009.

[61] A. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002.

[62] R. Mazza and A. Berr, "Focus group methodology for evaluating information visualization techniques and tools," *11th International Conference Information Visualization*, pp. 74–80, 2007.

[63] Z. L. X. L. M. Z. Mingli Song, Dacheng Tao, "Image ratio features for facial expression recognition application," 2009.

[64] R. Ohbuchi, K. Osada, T. Furuya, and T. Banno, "Salient local features for shape-based 3d model retrieval," pp. 93–102, 2008.

[65] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *IEEE International Conference on Multimedia and Expo*, 2005.

[66] Y. Shen, D. Chen, X. Tian, and M. Ouhyoung, "3d model search engine based on lightfield descriptors," *Eurographics*, 2003.

[67] X. Tang and Z. Li, "Audio-guided video-based face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 955–964, 2009.

[68] Y.-L. Tian, T. Kanade, and J. Cohn, "Recognizing lower face action units for facial expression analysis," in *4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, 1999, pp. 484–490.

[69] C. Wang and F. Wang, "A knowledge-based strategy for object recognition and reconstruction," pp. 387 – 391, 2009.

[70] J. Yang, D. Zhang, A. Frangi, and J.-y. Yang, "Two-dimensional pca: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, 2004.

[71] U. Yang, H. Shin, and K. Sohn, "Pose and illumination invariant 2d to 3d facial recognition system," 2008.

[72] M. Yeasin, B. Bullot, and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," *Ieee Transactions on Multimedia*, vol. 8, no. 3, pp. 500–508, 2006.

[73] Z. Yongmian, J. Qiang, Z. Zhiwei, and Y. Beifang, "Dynamic facial expression analysis and synthesis with mpeg-4 facial animation parameters," *IEEE Transactions on Circuits; Systems for Video Technology*, vol. 18, no. 10, pp. 1383–1396, 2008.

[74] Z. Zhang, "Feature-based facial expression recognition: Sensitivity analysis and experiments with a multi-layer perceptron," *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 1998.

[75] Q. Zhao, S. Xu, Z. Z. Li, and L. Wang, "A comparative study of musical navigation methods for visually impaired users of gui systems," in *2007 IEEE IEEM*, IEEE, Ed., vol. 1-4244-1529-2/07, 2007.