Expression: A Dyadic Conversation Aid using Google Glass for People who are Blind or Visually Impaired

ASM Iftekhar Anam, Shahinur Alam, and Mohammed Yeasin
Dept. of Electrical and Computer Engineering
University of Memphis
Memphis, TN 38152
Email: {aanam, salam, myeasin}@memphis.edu

***Abstract*—This paper presents 'Expression' — an integrated assistive solution using Google Glass. The key function of the system is to enable the user in perceiving social signals during a natural dyadic conversation. The design and implementation of the system addressed a number of technical and research challenges — video acquisition and communication over Wi-Fi, efficient detection and tracking of faces, overheating of Google Glass, robust detection of facial features and modeling behavioral expressions, and feedback system for perceiving social signals. Performance evaluation was conducted to ensure the completeness and generalizability of models. Furthermore, usability studies were performed with ten (10) subjects (six visually impaired and four blind-folded) to illustrate the utility of the 'Expression'. Subjective evaluation of Expression was performed using a five (5) point Likert Scale and was found to be excellent (4.383).**

I. Introduction

Limited access to non-verbal communication cues hinders the dyadic conversation or social interaction of people who are blind or visually impaired. Studies (e.g. [1]) have shown that social signals are often communicated through nonverbal channel. In a number of experiments, Argyle et al. [2] reported that nonverbal cues play a significant role in rating friendly or hostile attitude. Interviews with people who are blind or visually impaired revealed that they are more interested about the interlocutor's appearance, facial features and behavioral expressions. There is also a consensus that technological advances are yet to overcome the barriers that limit their abilities to gain independence. Assistive technology solutions can potentially help them from social isolation, lack of employment, depression, and other mental health issues.

In this paper, we present an integrated system, called **Expression** using Google Glass (see figure 1). The Google Glass (henceforth termed as the Glass) has an Optical Head Mounted Display designed to be worn as eyeglasses and is equipped with a camera, voice recognition, Internet connectivity, and an array of sensors. Such a system offers new possibilities, for example voice commands, Web search, hands-free interaction with smart phones, etc. However, due to small form factor, the battery life and heat dissipation of the device are not suitable for continuous usage and intensive computation. Hence, intensive computations such as detection and modeling of behavioral expressions were delegated to a dedicated server to avoid overheating of the Glass.

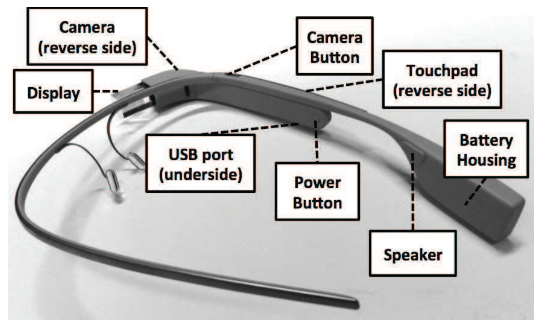

Fig. 1. Components of the Google Glass [3]

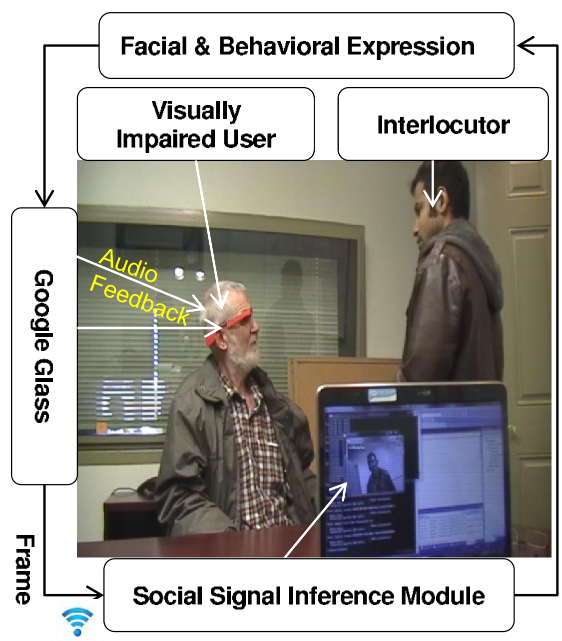

Fig. 2. Experimental Setup for evaluation of the **Expression**

The integrated system is designed to perceive social signals and provide feedback in soft real-time in an unobtrusive manner. Facial appearance features, behavioral expressions, body postures, and emotions are considered social signals

MobiCASE 2014, November 06-07, Austin, United States

DOI 10.4108/icst.mobicase.2014.257780

in this context of dyadic conversation. Vinciarelli et al [4] provides a detailed introduction, the challenges and potential implications of social signal processing. Figure 2 depicts the use case scenario of **Expression**. A blind or low vision user wears a Glass with the **Expression** application installed. The application captures video stream (5 -10 Frames per second) using the Glass camera and transmits to the server. The server analyzes the facial image and returns the detected facial features or behavioral expression. Speech feedback (such as 'Smile', 'Looking Right', 'Yawn' etc.) is conveyed to the user through ear bud using built-in Text-to-Speech service.

The implementation of the **Expression** followed the fusion of "participatory design" and "design thinking". The functionality of the system evolved over time through meaningful interaction with representative users[1]. The ideas of *empathy* for the context of a problem, *creativity* in the generation of insights and solutions, and *rationality* in analyzing and fitting various solutions to the problem were adopted from "design thinking" [5]. It is the fusion of ideas that lead us to use Google Glass as a platform to render services to people who are blind and visually impaired.

During the participatory design phase we interviewed and collaborated with three subjects to finalize the functionality of **Expression**. To detect the facial features, and model behavioral expressions, we collected dyadic conversation data from twenty (20) subjects both sighted and visually impaired. We annotated the videos to create a "ground truth" data to model social signals. The models were evaluated using recorded evaluation sessions (overall F-measure is 0.773). Usability study using 10 subjects (six blind and low-vision and four blind-folded) was performed in a five point Likert Scale.

The design and implementation of the **Expression** addressed a number of technical and research challenges — participatory design to understand users' need and system specification, video acquisition and communication over Wi-Fi, efficient detection and tracking of faces, overheating of the Glass, robust detection of facial features and modeling behavioral expressions, and feedback system for perceiving social signals.

In summary, the key contributions of the papers are:

- to build a framework for data annotation and stratification;
- real-time detection and tracking of faces in natural dyadic conversation;
- robust inference of social signals in soft real-time in an unobtrusive manner;
- building a fully integrated assistive solution for people who are blind and visually impaired to facilitate dyadic conversation;
- comprehensive usability study to demonstrate the utility of **Expression**.

[1]By representative users we refer to the people who are blind or visually impaired or low-vision.

II. Related Works

A plethora of assistive solutions have been developed to aid the people who are blind visually impaired. Velazquez [6] compiled a comprehensive list of wearable assistive devices worn in different body areas such as fingers, wrist, arm, abdomen, chest, head, feet, tongue, ear etc. Luo et al [7] developed a wearable vision enhancement device based on head mounted display (HMD) to aid the people with vision problem. The system supports alternating see through and magnification mode to assist the people with disability in seeing distant objects and signs.

Dristi [8] is a wearable system for the blind to navigate in indoor and outdoor environment. It packaged a portable computer in a backpack for data processing from the beacons attached to the body of the user. Krishna et al.[9] developed a prototype of social interaction assistant to facilitate learning and recognizing faces. A number of the wearable systems for the people with disability were developed for navigational aid. [10] provided an experimental analysis of a sign-based way finding system using a cellphone camera. It detects "landmarks" in the environment (e.g. Room number and Name of occupant in an office) and guides a person without sight towards the detected landmark.

With the introduction of iPhone in 2007, there started a new wave of visual aids in the form of smartphone applications. As the Glass came out in 2013, there is another noticeable shift towards wearable assistive solutions. Shilkrot and colleagues [11] developed a wearable device that assists the visually impaired in reading printed texts. Juan and colleagues [12] see the advent of wearable computing platform as the beginning of a new generation of hands-free assistive vision applications that would render seamless experience in social interactions.

The "Team F.A.C.E." [13] developed a facial and expression recognition system for the blind and low-vision people mounted on a standard white cane. It detects six basic emotions defined by Ekman [14] whereas we argue that though these emotions are prevalent in general social interactions, in dyadic conversations the emotions are more nuanced. Hence, we focus on facial and behavioral expressions instead of categorical emotions. Gade et al [15] proposed a social interaction assistant that uses a wearable camera to localize persons. Most relevant to our work is ***i*MAPS** [16] — a smart phone based prototype that predicts affective dimensions (valence, arousal, and dominance) in social interactions. Another similar work is ***i*FEPS** [17], a sensory substitution system that produces auditory feedback for changes in facial expressions for the users. Both the systems are implemented on a smartphone platform and have their limitations in terms of deployment such as hanging the phone from the neck, limited field of view of the phone camera, etc. A number of challenges arose due to different form factor of the Glass — especially the heating problem, short battery life, and bandwidth for data transmission. Additionally, we collected and annotated data from both sighted and visually impaired individuals to model the social signals in a natural setting. It is important to note that the interlocutor can be either sighted or have disabilities.

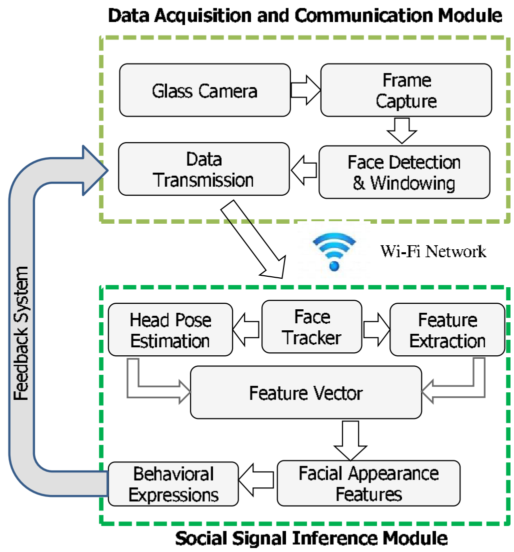

Fig. 3. Diagram of the complete Expression System

III. Design of 'Expression'

Design and implementation of **Expression** followed the ideas from participatory design and went through a number of iterations. Figure 3 shows the system architecture of **Expression**. It has three modules:

- data acquisition and communication
- social signal inference module
- feedback system.

The rationale of the system design has been discussed in the following sections. The data acquisition module captures video stream using the Glass camera. It uses a Viola-Jones face detector from OpenCV library for fast detection of face region in camera preview mode. A region of interest (ROI) around the face location is selected by windowing technique. After JPEG compression, the data is transmitted to the server. The server runs the social signal inference module that extracts facial features and estimates head pose to generate feature vector. A rule based classifier infers the behavioral expressions from feature vectors using a time based sliding window technique. The output is then sent to the feedback system where built-in Text-to-Speech service generates the audio feedback.

A. Design Activities in the Lab

We started our design process to address the following research questions:

- What are the behavioral expressions in the context of dyadic conversations?
- How to capture a balanced dataset to model various behavioral and facial expressions?
- How to select the features to model the expressions?

The following subsections describe our design activities.

1) Selection of Behavioral Expressions:: **Expression** was designed to help perceiving social signals in a natural dyadic conversation. In particular, the system was designed to assist the blind or visually impaired. As a part of the design, we studied psychology literature to understand facial features and behavioral expressions for social signals. Duncan [18] reported a set of body motions from his research on communication behaviors in face-to-face interaction:

- head gestures and movement (nodding, turning, pointing, shaking, etc.);
- shoulder movements (e.g. shrugs);
- facial expressions;
- hand gestures, movements, and different hand positions;
- arm movements and positions;
- foot movements;
- leg movements and positions;
- postures and shift of postures;
- use of artifacts such as pipe, papers, and clip board.

In a recent study of dyadic conversation, Cummins [19] studied the role of gaze and eye blink in the context of conversation and turn taking. Kleinke [20] explored the gaze and making eye contacts during conversations as a means to provide information, regulate interaction, express intimacy, and exercise social control. Patterson reported interpersonal distance, gaze, facial expressiveness, and head nods as the constructs of nonverbal involvement [21].

The main challenge of modeling behavioral expressions are data collection, annotation, feature selection, and modeling. We collected a set of behavioral expressions through user study:

- head movements (look up/down, look left/right, tilt left/right);
- facial expressions (smile, open smile); and
- behavioral expressions (yawn, sleepy).

Figure 4 shows a few examples of the selected expressions.

B. Data Collection and Annotation

The existing datasets in facial expression and emotion research are mostly built by emotion elicitation techniques and included only sighted people. Since the visually impaired individuals often interact with both sighted and blind or low vision people, we decided to include the visually impaired individuals in our data collection along with the sighted people. The authors played the role of the interviewers and the participants were asked to engage in a dyadic conversation about topics of their interest. Each conversation lasted for about 10 minutes and was recorded using the Glass camera worn by the interviewer. Six (6) visually impaired and 14 sighted people participated in the data collection process.

Five annotators performed frame by frame annotation of the data. The authors demonstrated example annotation tasks to train them. The inter-rater agreement measured by Fleiss'

Fig. 4. Example of expressions detected(from left): Tilt Left, Looking Right, Tilt Right, Smile, Looking Up, Yawn, Looking Down, Looking Left

Kappa [22] was 0.791 that indicates significant agreement. Though it can't be generalized across the population due to small number of participants, the annotators observed that the congenitally blind subjects were less expressive compared to their non-congenital counterparts. After cleaning up the noises from recordings, we obtained about 2 hours of dyadic conversation data.

C. Feature Selection and Model Training

The Facial Action Coding System (FACS) [23] represents facial expressions as the movements of Action Units (AUs). Rahman et al [24] proposed a framework that captures the relationships among the *Frequently Occurred and Strongly Connected* AUs and predicts the relations among *Infrequently Occurred and Weakly Connected* ones. Based on the study, we select facial features that are correlated to most of the AUs and, hence, will be able to include more facial expressions. We used distance based features such as

- height of inner eyebrow and outer eyebrow;
- height of eye opening;
- height of inner and outer lip boundary; and
- distance between lip corners.

Facial features and head pose are extracted using a Constrained Local Model (CLM) based face-tracker. Developed by Saragih et al [25] it fits a parameterized 3D shape model and track 66 facial landmark points in the image. A canonical reference shape was obtained from the mean of the shapes of different expressions. The features were calculated as the ratio of the distances between appropriate landmarks after removal of global transformation. The head pose was estimated from the tracked 3D shape. We then trained the rule-based classifiers using the annotated data.

D. Integration of Google Glass and Server

This subsection describes the development platform for Google Glass and our implementation details of **Expression**.

1) Google Glass Development Platform: Google Glass applications (also known as Glassware) can be developed in two different ways – either using official Mirror API or Glass Development Kit or GDK. Mirror API is a cloud service that communicates with the applications via RESTful messages. The contents can be viewed as time-line cards on the device display. The cards use any of the predefined layouts or a custom layout based on a limited subset of HTML. In order to ensure real-time interaction it requires stable Internet connection. Also, the API does not allow modifying Glass resolutions and display settings. On the other hand, GDK is an add-on of the existing Android platform and supports access to all the hardware sensors. Therefore for the applications that require real-time responsiveness, GDK is the preferred choice.

2) Prototyping of 'Expresssion': The Google Glass was chosen as the development platform for a number of reasons. It is more ergonomic than a head mounted or neck mounted camera often used in prototyping assistive vision systems. A blind person does not look any different from a sighted person wearing the Glass. Therefore it is less conspicuous which they prefer when in public places. Moreover, the technical features of the Glass made it an ideal choice to develop the system that requires video capture, streaming, and speech feedback. For our application, we are not using the head mounted display for any kind of feedback. Due to short battery life, the application cannot run for an extended time on the Glass. However, our purpose was to demonstrate how such a wearable device could be useful for the people with limited or no vision. It captures video stream and transmits to the server. The server processes the frames, tracks the facial features, and sends the detected expressions back to the Glass application. The application then generates audio feedback using Text-to-Speech engine. We adopted the audio feedback according to the user study conducted to evaluate the ***i*FEPS** [17] system.

E. Participatory Design

For the next phase of system improvement, we performed the participatory design through collaboration with a small group of representative users. Clovernook Center for the Blind and Visually Impaired and the Mid-South Access Center for Technology (Mid-South ACT), both located in Memphis, are the two venues where we conducted our research. Clovernook Center is especially for the blind and visually impaired individuals and provides training with assistive technology devices. Mid-South ACT is a division of Center for Rehabilitation and Employment Research (CRER) and provide training for clients with different disabilities. It also provides systematic evaluation of assistive technology and conducts research and outreach program.

1) Participants: We evaluated the **Expression** system with a total of 10 participants and 6 of them were blind or low vision. The participants were recruited through contacting the directors of the centers and student disability services at the University of Memphis. Among the 6 visually impaired participants (4 female) 4 were African-American, one Caucasian, and one Asian. Our goal was to include participants from different ethnicity, age, and gender. They had age range between 29 and 65. The sighted participants were all graduate students from the department of Electrical and Computer Engineering at the University of Memphis. Their age range was between 25 and 30. Table I lists the details of the visually impaired participants.

The focus group consisted of three representative users with varying degrees of disability (P1, P2, and P6 in Table I). One of them is congenitally blind in one eye and gradually lost vision entirely, the other participant lost vision before reaching teen age, and the third participant lost vision after the age of 40 due to diabetes and stroke. The individuals were selected based on their availability, experience, interest, and familiarity with technology. They participated in in-person interviews, brainstorming sessions, and phone interviews and suggested modifications and improvements. They evaluated the the prototype of **Expression** and we modified the application based on their recommendations and suggestions. In the "Technology Use" column of Table I, **Extensive** refers to the users who are adept in using various assistive technology and computer programs such as JAWS$^{\text{TM}}$, VoiceOver$^{\text{TM}}$, ZoomText$^{\text{TM}}$etc. They use smart phones extensively and manages a good set of assistive applications for regular use. **Moderate** users are familiar with smart phones and computer software but have not mastered the use, and the **Low** exposure users are novice or use technology rarely.

2) Evolution through Design Thinking: Google Glass is not suitable for continuous mobile vision applications since the battery drains quickly and the device heats up if the camera is continuously used even for a short duration. Likamawa et al [3] investigated various use cases of the Glass to quantify the power consumption and characterize temperature profile. The alpha version of **Expression** application on Glass only gave feedback when it got the result from server. The users found it confusing when there was no face on the screen or face tracker failed on the server due to small face size or out of plane head movements. If the server can't track any face on the frames, it sends "No face found" feedback. However, it was not timely and the users were not satisfied with that.

To accommodate their need, we incorporated a Viola-Jones face detector from OpenCV library for Android to select a region of interest (ROI) containing the face [26]. It serves three functions:

- we can provide feedback to the user about the position of the face on the screen so that they can adjust their postures accordingly;
- we transmit only the ROI to the server to optimize data transmission;
- we also provide feedback to the user to move close towards the interlocutor when the face size is smaller than 40×40.

After detecting a face, the size of a bounding box is calculated. Then a rectangle is created whose dimension is double the size of the face bounding box. If the dimension exceeds the image boundary, which occurs when face is detected farther from the center of the frame, the boundary of ROI is adjusted with respect to the frame. Figure 5 and 6 show two possible scenarios of windowing.

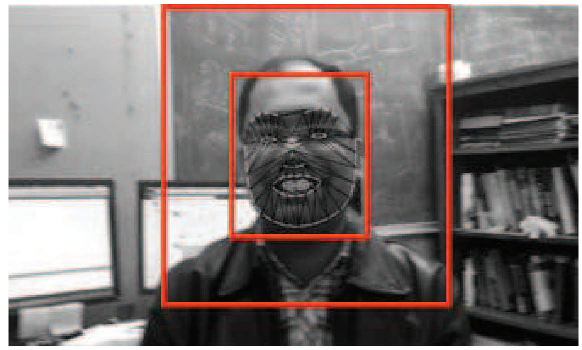

Fig. 5. Windowing in a frame where face is centered

Fig. 6. Windowing in a frame where face is at the corner in the frame

Besides optimizing data transmission, selecting a ROI around a face also have the following advantages:

- The face tracker fails rarely as it finds a face in all the frames and
- Due to large face size, feature extraction is also optimized and it improves the response time of the server.

However, adding the face detector to the application exacerbated the heating problem. In another work Likamawa and colleagues [27] addressed optimization of energy usage for continuous mobile vision. They showed that with the existing mobile camera sensors it is possible to achieve constant energy per pixel for video capture at low frame rates. Therefore, we downgraded the standard frame rate of the Google Glass. We empirically set the minimum frame rate to 5 FPS and maximum to 10 FPS.

3) Design of Feedback: Feedback in an assistive solution is an important design consideration influenced by personal preferences, surroundings, and context of use. From the interviews it was evident that speech feedback is the most desired mode. It is also supported by the findings from the study in [17]. A speech based feedback was designed for the **Expression**. In the initial **Expression** system, the social signal inference module continuously spotted the behavioral expressions. When evaluated by the participatory design team, they suggested to generate feedback when there is change in expressions instead of continuous output. The feedback

TABLE I. STATISTICS OF BLINDNESS AND TECHNOLOGY EXPERTISE OF THE PARTICIPANTS

ID	Age	Gender	Race	Nature of Impairment	Technology Use
P1	$61-65$	M	Caucasian	Congenitally Blind	Extensive
P2	$36-40$	M	Asian	Blind since teenage	Extensive
P3	$56-60$	F	African-American	Partial vision on left eye	Moderate
P4	$26-30$	F	African-American	Partial vision on left eye	Moderate
P5	$26-30$	F	African-American	Partial vision in bright light	Low
P6	$56-60$	F	African-American	Partial vision on left eye	Extensive

was then modified to keep track of the expressions, and we give feedback only when a new expression is detected or the previous expression sustains for a longer duration. It is easy to follow the conversation with the modified feedback.

IV. EVALUATIONS AND RESULTS

We conducted both qualitative and quantitative evaluations for **Expression** system. To evaluate **Expression**, we asked the subjects to engage in two dyadic conversations with the interviewer. Each of the conversations lasted for about 10 minutes. In the first session they did not wear the Glass, and in the second conversation they put the Glass on with the **Expression** application installed. They talked about the topics of their interests collected through a set of questionnaire. 10 subjects (six visually impaired and four blind-folded sighted persons) participated in the study conducted at the MidSouth ACT and the Clovernook Center. We report the results in the following sections.

A. Quantitative Evaluation of 'Expression'

We recorded and annotated the dyadic conversation sessions to evaluate the system performance. The annotators annotated the recordings and examined the speech feedback produced by the **Expression**. The precision (or positive predictive value) and the recall (sensitivity or true positive rate) values of different behavioral expressions are shown in Table II. The F-measure (or F_1 Score) is calculated from the average values of the precision and recall. The overall F-measure is 0.773 which is reasonable for a real-time systems such as the **Expression**. Since dyadic interaction literature contains many other expressions, it would be interesting to see how the system performs when more expressions are included to the vocabulary which we leave as a future work.

B. Qualitative Evaluation

We followed up with the participants who were blind or low vision for subjective evaluation after the dyadic conversation sessions using **Expression**. We asked the participants to rate their confidence with a set of statements in a 5-point Likert Scale with 5 being the highest. The statements were about correctness, learnability, informativeness, usability, portability, and user satisfaction after using the system. We report the result of the subjective evaluation in Figure 7

We also asked their opinions about positive and negative aspects of the **Expression** and any issues related to Google Glass. They mentioned a number of positive aspects such as:

- hands-free interaction,
- tracking subtle changes,
- lightweight device, and

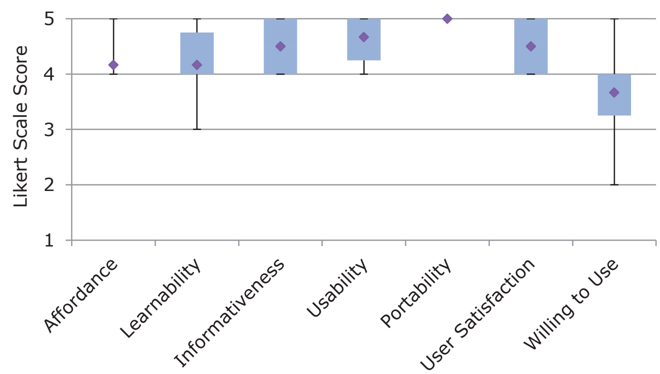

Fig. 7. Result of subjective evaluation (higher score is better)

- speech feedback of the expressions instead of tones.

As the limitation of the system, one participant complained about sporadic short delays in the feedback. Also, being partially sighted, she wanted the frame not to occupy the full screen so that she can use her eyesight. Another participant expected the Glass to be a little bigger in size. To answer the question of "Would you feel comfortable using such a system in public?", one participant replied:

```
Google Glass is less alienating compared
to white cane. When I go out with my cane
people runs to the hill to leave room for
me. But I hope they would not notice this
[Google Glass] from a distance. Also the
benefits would eventually outweigh the
concerns and [I hope] it will be more
common.
```

V. DISCUSSION & CONCLUSION

We present **Expression** system that evolved through the participatory design approach and design thinking. The implementation and a thorough evaluation substantiated the effectiveness of the system. We describe the insights obtained through the process and suggest outstanding issues for future development and improvement of the user experience with wearable devices. With that said, our work is not beyond limitations. Here we briefly describe the limitations that can affect the functioning of the system.

Multimodal feedback: In our study, a low vision participant wanted tones along with the speech feedback. Her opinion was if she misses the speech feedback while concentrating on the conversation, tones will be helpful. The idea of multimodal feedback seemed interesting to the participatory design team and is currently work in progress.

TABLE II. PRECISION AND RECALL OF THE EXPRESSIONS

Expressions	Total Events	Predicted	Missed	False Alarm	Precision	Recall
Smile	23	21	2	6	0.913	0.778
OpenSmile	18	15	3	3	0.833	0.833
Sleepy	12	9	3	4	0.75	0.692
Yawn	8	5	3	2	0.625	0.714
Looking up/down	19	17	2	7	0.895	0.708
Looking left/right	15	14	1	8	0.933	0.636
				Average Precision & Recall	**0.825**	**0.727**

TABLE III. USABILITY TEST STATEMENTS

Criteria	Statement [5 - Strongly Agree and 0 - Strongly Disagree]
Affordance	**Expression** can be successfully used to understand others' facial expressions
Learnability	**Expression** can be successfully used to understand others' emotion in social context
Informativeness	**Expression** conveys more information about a person's face than it is naturally expressed by a person's voice
Usability	**Expression** is easy to use in daily life
Portability	**Expression** is portable
User satisfaction	**Expression** improves my social interactions with the interviewer
Willing to Use	I will use **Expression** in my daily life

Extending vocabulary of expressions: We resorted to a set of common facial and behavioral expressions to develop the dataset for the **Expression**. Building a comprehensive data set with proper annotation is quite challenging. We plan to add more expressions (one participant suggested frowning), head movement and hand gestures which are quite common in dyadic conversations.

Extraneous head movements: During the annotation of the videos collected using Google Glass we found that some segments of videos contain extraneous head movements of the interviewer and the tracker often failed to track the face. We discarded those segments since it was beyond the scope of our current work.

Detecting Eye Blink: Eye blink is an interesting expression to be included in the vocabulary. It is a rapid event that requires high frame rate data acquisition for successful detection. There are works on detecting eye blink [28] using USB cameras. However, implementing the solution in Google Glass is difficult due to limited computing and battery resources.

ACKNOWLEDGMENTS

We are thankful to Dr. Lavonnie Claybon, Director of Midsouth Access Center for Technology to provide access to representative subjects to collect data and evaluate **Expression**. We thank the participants for their feedback in the design and evaluation. This work was partially funded by National Science Foundation (NSF–IIS-0746790), USA. Any opinions, findings, and conclusions or recommendations do not reflect the views of the funding institution.

REFERENCES

[1] P. M. Brunet, H. Donnan, G. McKeown, E. Douglas-Cowie, and R. Cowie, "Social signal processing: What are the relevant variables? and in what ways do they relate?" in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 2009, pp. 1–6.

[2] M. Argyle, F. Alkema, and R. Gilmour, "The communication of friendly and hostile attitudes by verbal and non-verbal signals," *European Journal of Social Psychology*, vol. 1, no. 3, pp. 385–402, 1971.

[3] R. LiKamWa, Z. Wang, A. Carroll, F. X. Lin, and L. Zhong, "Draining our glass: An energy and heat characterization of google glass," *arXiv preprint arXiv:1404.1320*, 2014.

[4] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, no. 12, pp. 1743–1759, 2009.

[5] T. Brown *et al.*, "Design thinking," *Harvard business review*, vol. 86, no. 6, p. 84, 2008.

[6] R. Velázquez, "Wearable assistive devices for the blind," in *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment*. Springer, 2010, pp. 331–349.

[7] G. Luo, Z. Li, and E. Peli, "Mobile electronic magnification device for people with central vision loss," *Invest Ophthalmol Vis Sci*, vol. 52, 2011.

[8] L. Ran, S. Helal, and S. Moore, "Drishti: an integrated indoor/outdoor blind navigation system and service," in *Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on*. IEEE, 2004, pp. 23–30.

[9] S. Krishna, D. Colbry, J. Black, V. Balasubramanian, S. Panchanathan *et al.*, "A systematic requirements analysis and development of an assistive device to enhance the social interaction of people who are blind or visually impaired," in *Workshop on Computer Vision Applications for the Visually Impaired*, 2008.

[10] R. Manduchi, "Mobile vision as assistive technology for the blind: An experimental study," in *Computers Helping People with Special Needs*. Springer, 2012, pp. 9–16.

[11] J. H. Roy Shilkrot, P. M. Connie K. Liu, and S. Nanayakkara, "Fingerreader: A wearable device to support text-reading on the go," in *In Proceedings of CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2014.

[12] J. R. Terven, J. Salas, and B. Raducanu, "New opportunities for computer vision-based assistive technology systems for the visually impaired," *Computer*, vol. 47, no. 4, pp. 52–58, 2014.

[13] D. Astler, H. Chau, K. Hsu, A. Hua, A. Kannan, L. Lei, M. Nathanson, E. Paryavi, M. Rosen, H. Unno *et al.*, "Increased accessibility to nonverbal communication through facial and expression recognition technologies for blind/visually impaired subjects," in *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2011, pp. 259–260.

[14] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.

[15] L. Gade, S. Krishna, and S. Panchanathan, "Person localization using a wearable camera towards enhancing social interactions for individuals with visual impairment," in *Proceedings of the 1st ACM SIGMM international workshop on Media studies and implementations that help improving access to disabled users*. ACM, 2009, pp. 53–62.

[16] A. Rahman, M. Tanveer, A. Anam, and M. Yeasin, "Imaps: A smart phone based real-time framework for prediction of affect in natural dyadic conversation," in *Visual Communications and Image Processing (VCIP), 2012 IEEE*, Nov 2012, pp. 1–6.

[17] M. I. Tanveer, A. Anam, M. Yeasin, and M. Khan, "Do you see what I see?: designing a sensory substitution device to access non-verbal

modes of communication," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 2013, p. 10.

[18] S. Duncan Jr, "Toward a grammar for dyadic conversation," *Semiotica*, vol. 9, no. 1, pp. 29–46, 1973.

[19] F. Cummins, "Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals," *Language and Cognitive Processes*, vol. 27, no. 10, pp. 1525–1549, 2012.

[20] C. L. Kleinke, "Gaze and eye contact: a research review." *Psychological bulletin*, vol. 100, no. 1, p. 78, 1986.

[21] M. L. Patterson, "A sequential functional model of nonverbal exchange." *Psychological Review*, vol. 89, no. 3, p. 231, 1982.

[22] J. L. Fleiss, B. Levin, and M. C. Paik, "The measurement of interrater agreement," *Statistical methods for rates and proportions*, vol. 2, pp. 212–236, 1981.

[23] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.

[24] A. M. Rahman, M. I. Tanveer, and M. Yeasin, "A spatio-temporal probabilistic framework for dividing and predicting facial action units," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 598–607.

[25] J. M. Saragih, S. Lucey, and J. F. Cohn, "Face alignment through subspace constrained mean-shifts," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1034–1041.

[26] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[27] R. LiKamWa, B. Priyantha, M. Philipose, L. Zhong, and P. Bahl, "Energy characterization and optimization of image sensing toward continuous mobile vision," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 69–82.

[28] M. Chau and M. Betke, "Real time eye tracking and blink detection with usb cameras," Boston University Computer Science Department, Tech. Rep., 2005.