LITTLEHELPER: AN AUGMENTED REALITY GLASS APPLICATION TO ASSIST INDIVIDUALS WITH AUTISM IN JOB INTERVIEW*

Qingguo Xu[1], Sen-ching Samson Cheung[1], and Neelkamal Soares[2]

[1]Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY 40506
qingguo.xu@uky.edu, sccheung@ieee.org
[2]Autism & Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA
nssoares@geisinger.edu

ABSTRACT

With the rapid increase in the prevalence of autism spectrum disorder (ASD) in 1990s, there are approximately 50,000 individuals with ASD turning 18 years old every year. The community-based employment rate, even for those individuals with higher functioning capabilities, is very low. This can be partly explained by socio-communicative skill deficits which are a hallmark of ASD, such as poor eye-contact and inappropriately modulated speech. There has been little work in developing assistive technology to help individuals with ASD to compensate for these deficits. In this paper, we propose a new system based on a wearable augmented-reality glass platform called LittleHelper to provide customized supports for individuals with ASD in enhancing social communication during a job interview. Using the built-in camera and microphone, LittleHelper can detect the position of the interviewer relative to the center of the camera view, and measure the sound level of the user. Based on these inputs, appropriate visual feedbacks are provided back to the user through the optical head-mounted display.

***Index Terms*—**Autism Spectrum Disorder (ASD), Augmented Reality Glass, Job Interview, Speech Perception, Head Tracking

1. INTRODUCTION

There has been a rapid increase in the prevalence of autism spectrum disorder (ASD) since the 1990s [1], and there are approximately 50,000 individuals with ASD turning 18 years old every year [2]. Competitive employment upon graduation is a critical outcome for all individuals and has been correlated with life satisfaction, better health, and increased social outcomes. Unfortunately, data suggest that the majority of individuals with ASD have difficulties obtaining and maintaining competitive employment upon high school graduation. Several research teams have reported disproportionately low rates of employment for individuals with ASD [2], indicating that regardless of intellectual functioning, they are less likely to be employed than their typically developing peers and individuals with learning disabilities.

The hallmark socio-communicative impairments of ASD, including poorly modulated eye-contact and inappropriately modulated speech, could be barriers to competitive employment. These impairments could provide an inaccurate impression of one's ability during a job interview to employers who are not familiar with ASD. The employer may feel offended if the interviewee keeps looking away while the employer is talking to him/her. In addition, the inappropriately stressed and monotonous speech of an individual with ASD may affect the expressivity and effectiveness of the conversation [3].

There has been little work done previously in developing assistive technology to help individuals with ASD compensate for these deficits. One study team proposed a training system for individuals with ASD to learn how they should behave during a job interview [4]. For such a training system to be effective, an individual must be self-conscious of his/her behavior and mental state so as to make the appropriate adjustment in real-time during a social interaction. These skills are conceptualized as being part of the Theory of Mind, which is also believed to be impaired in ASD [5]. While long-term behavioral intervention has been shown to be effective in improving self-awareness [6], assistive technology has a role in providing feedback, such as visual feedback, for self-awareness training to individuals with ASD.

The recent advancements in wearable technologies from body sensors to augmented-reality glasses provide an opportunity to deploy and test such technologies. One such example, Google Glass, includes a camera, microphone, optical head-mounted display (OHMD) and other components to support wearable computing and networking capability. The camera can record live video from the perspective of the user, and the microphone can capture both the user's voice and ambient noise. These provide the sensing signals of the user's environment. Unlike the display of a typical handheld mobile device, the OHMD is a smaller, transparent display that provide unobtrusive visual feedback. Google Glass can guide a subject in reacting to environmental changes by geometrically aligning the visual cues with the environment. There is prior work in using Google Glass in navigation for

* Part of this material is based upon work supported by the National Science Foundation under Grant No. 1237134.

visually impaired individuals [7] and emotion recognition [8]. To the best of knowledge, there is no prior work utilizing visual feedback to address eye contact and speech stressing using Google Glass.

In this paper, we propose a Google Glass application, LittleHelper, to assist individuals with ASD in maintaining appropriate eye contact and speech volume in a one-on-one interaction. In a broader sense, we intend this application as an aid for individuals during a typical job interview process. LittleHelper includes two components: the speech volume detection module which helps users to be more aware of the loudness of their voice, and the interviewer detection module which tracks where the interviewer is and provide simple visual cues to establish appropriate gaze.

The remaining part of this paper is organized as follow: We describe the design details in section II. Section III includes experimental results and a comparison with using built-in sensors to track the relative movement. Section IV concludes this paper.

2. DESIGN OF LITTLEHELPER

LittleHelper consists of two main processing components: speech volume level detection and interviewer detection. Figure 1 shows the graphical user interface as seen by the user on the OHMD. It contains the real-time view of the forward facing camera with a "progress bar" at the top left corner. The progress bar indicates the volume of the microphone input, and will alert the user if the speaking voice becomes either too loud or too soft. The camera view provides visual feedback on the face of the interviewer. When the face is off center, an arrow will be shown to direct the user's head pose so as to reestablish proper eye-gaze.

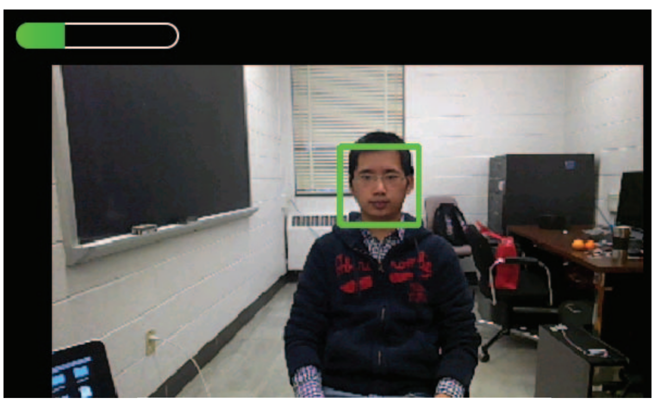

Figure 1: GUI of LittleHelper

While our prototype is still in alpha stage and no clinical testing has been done, we have spent significant effort in incorporating domain experts and providing a design that assists individuals with autism without further distracting them from social communication. One of the co-authors, Dr. Neelkamal Soares, is a developmental pediatrician specialized in treating patients with autism. Our earlier prototype has also been accepted as a technical demonstration at the International Meeting for Autism Research (IMFAR) [12], during which we have collected expert opinions from autism experts from different countries to further improve our design.

2.1 Speech volume level detection

A key challenge facing the development of wearable real-time signal processing is the limitation in computation power of the platform. To reduce computing load in volume estimation, we set the sample rate to 8000 Hz and sample width to 16 bits, the smallest supported by the Android operating system. Using a window size of 1 second, we calculate the root mean square (RMS) of the audio signal $x(t)$, as defined in Equation (1), of the entire window W as a loudness measurement. The progress bar with a green-to-red gradient will show the amplitude in real-time.

$$RMS = \sqrt{\frac{1}{W}\sum_{t=1}^{W} x^2(t)} \quad (1)$$

In order for the measurement to be robust against ambient noise in the environment, we estimate an ambient noise floor level during the initial training phase when the user is not speaking. The appropriate speaking volume depends on the speech volume, the ambient noise level as well as the distance between the user and the interviewer. The first two are obtained through the microphone while the last quantity, the distance is estimated based on the size of the face as detected by the face detector. While the accuracy of this estimator depends on the head size of an individual, studies have shown that the variation in head size is relatively small among adults [13]. Our goal is to provide visual alert if the user's volume significantly deviates from this target "socially-acceptable" volume level. Figure 2 and 3 show how these visual feedbacks are displayed.

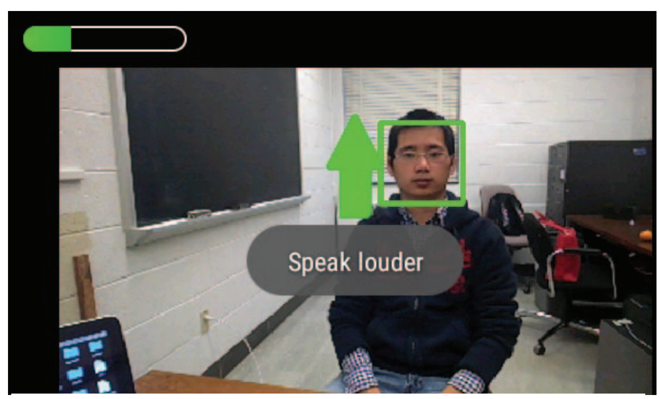

Figure 2: Feedback for speaking too soft

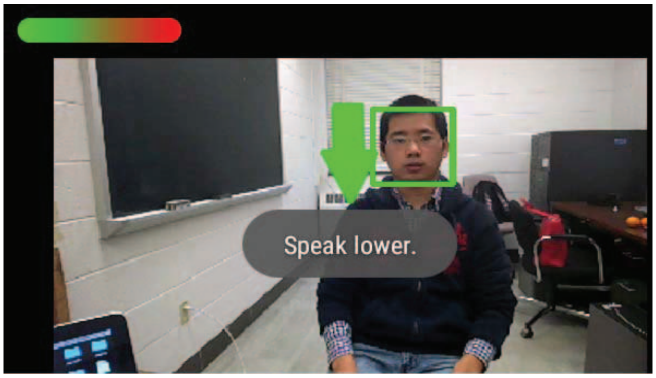

Figure 3: Feedback for speaking too loud

To determine the mapping from volume and distance to socially-accepted loudness level, we have conducted a human-subject experiment based on the opinion of five subjects, in response to the same speaker speaking at different volume levels at different distance from the subjects. The distance ranges from 0.5m to 3.0m at an interval of 0.5m. The detailed results of this experiment can be found in Section 5.

2.2 Interviewer detection

To detect the location of the interviewer's face, we rely on the Viola-Jones face detector [14] from the open-source OPENCV library adapted for Google Glass platform [9]. The face detection can run in real-time locally at the Google Glass. A common condition faced by individuals with ASD is a reduced capability in face processing, leading to poor eye-gaze in social communication. While eye gaze cannot be directly measured by Google glass, the head pose, as measured by the camera view, provides important cues to the visual attention. When the face is off-center, such as that shown in Figure 4, an arrow will appear at the center of the display to guide the user back to the interviewer.

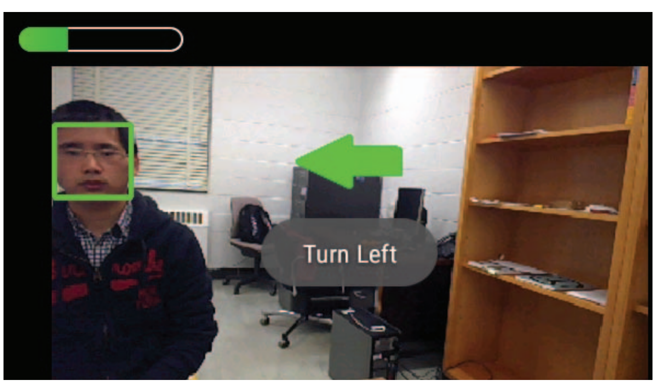

Figure 4: Feedback for interviewer disappearing from the view

The problem of relying solely on a face detector is that the interviewer may not be within the field of view of the camera. The field of the view of the Google Glass's camera is quite narrow (15 degrees horizontally). The interviewer can easily drop out of the visual field with a casual head movement from the user. Obviously, face detector alone will not be sufficient to provide the appropriate visual feedback to the user.

The Google Glass platform provides a total of seven sensors, including a 3-axis accelerometer and a gyroscope that can be used to measure the movement of the device [11]. Our initial testing with these two sensors shows that the measurements are quite noisy. They are useful in terms of measuring changes, which provide alternative means for human computer control to various functions. However, they are unsuitable to provide reasonably accurate tracking the actual head pose. Furthermore, there are two types of movement that we need to measure, the head pose movement and the relative movement of the interviewer with respect to the user. The second type of measurement cannot be accomplished based on movement sensors inside Google Glass.

As such, we have adopted an imaging solution to directly estimate the relative movement of the interviewer with respect to the user based on optical flow estimated from successive video frames. We estimate the global planar camera movement based the median of all the motion vectors, and the movement of the interviewer based on the movement of the facial area. Their difference provides an estimate of the relative position of the interviewer with respect to the user. If the facial region is out of the visual field, the last estimated position would be used and an arrow indicating the opposite direction will be displayed to guide the user back to the interviewer.

Due to the limited computing capability of the Google Glass, we are unable to run in real-time the optical flow algorithm on the platform. Instead, we stream the video through Wi-Fi to a local server and run optical flow program at the server. The measured relative position of the interviewer is sent back to the Glass for feedback rendering. An example of the estimated optical flow can be found in Figure 5.

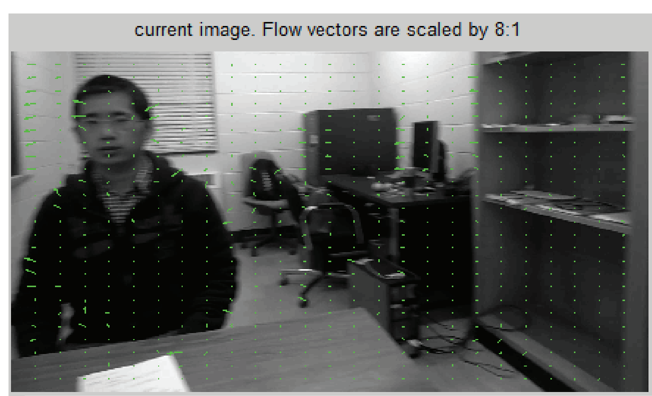

Figure 5: Estimated optical flow.

3. EXPERIMENTAL RESULTS

The Google Glass used in our experiments is the explorer edition with Android 4.4.4. The local server runs on a desktop (2 Quad CPU 3.00 GHz, 4GB RAM, 64-bit Windows 7 system) in the office. The optical flow program is from [10] implemented in MATLAB. With a 1-second audio buffer, the overall delay including the processing time is less than 1.5s, which is unnoticeable given the typical pace of movement and conversation in a job interview.

We have tested the proposed system in a small size office (4m×4m). The room is relatively quiet with the only noise source coming from the air conditioning unit. The measured ambient noise level is 5.384. Table I shows the relationship between the range of socially-accepted speech volume as a function of the distance between the user and the interviewer. The RMS range does not increase linearly with the distance – the increase of the acceptable volume levels off beyond 2.0m.

Table I: Relationship between distance and RMS value

Distance	RMS Value	Distance	RMS Value
0.5m	595~610	2.0m	710~725
1.0m	630~660	2.5m	730~740
1.5m	670~700	3.0m	745~760

Figure 6 shows the relationship between the distance and the face size. Once we detect the interviewer's face and obtain the face size, we can roughly estimate the distance between the user and the interviewer, and subsequently determine the socially accepted speech volume based on Table I.

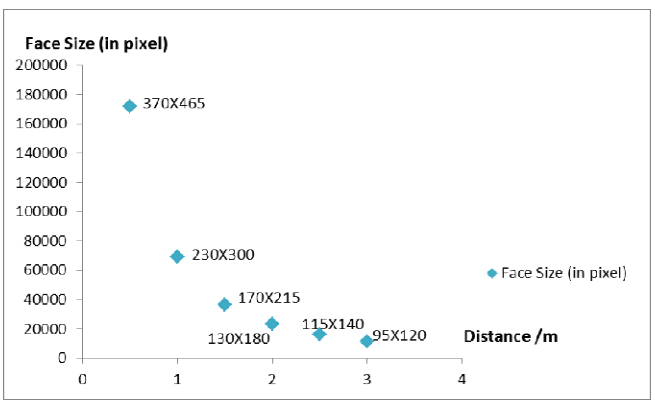

Figure 6: The relationship between face size and the distance

4. CONCLUSIONS

The advancement in wearable technology opens door to various assistive technology that can enhance social communication. In this paper, we have proposed a Google Glass application to help individuals with ASD to maintain proper eye contact and voice level during a job-interview-like setting. Our initial experiments have demonstrated an acceptable level of performance for the target application. We are currently in the process of recruiting human subjects with ASD to measure the usefulness of the system. The current technical limitations include the lack of tracking of true eye-gaze, the requirement of a server to offload complex computation and the inability to handle multiple persons in a social setting.

5. REFERENCES

[1] Gurney, J. G., Fritz, M. S., Ness, K. K., Sievers, P., Newschaffer, C. J., and Shapiro, E. G. "Analysis of prevalence trends of autism spectrum disorder in Minnesota," Archives of Pediatrics, vol. 157, pp. 622-627, 2003

[2] Shattuck, P. T., Narendorf, S. C., Cooper, B., Sterzing, P. R., Wagner, M., and Taylor, J. L. "Postsecondary education and employment among youth with an autism spectrum disorder," Pediatrics, vol. 129, pp. 1042-1049, 2012

[3] Russo, N., Larson, C., and Kraus, N. "Audio-vocal system regulation in children with autism spectrum disorders," Exp Brain Res, vol. 188, pp. 111-124, 2008

[4] Smith, M. J., Ginger, E. J., Wright, K., Wright, M. A., Taylor, J. L., Humm, L. B., Olsen, D. E., Bell, M. D., Fleming, M. F. "Virtual Reality Job Interview Training in Adults with Autism Spectrum Disorder," Journal of Autism and Developmental Disorder, vol. 44, pp. 2450-2463, 2014

[5] Baron-Cohen, Simon. "Mindblindness: An essay on autism and theory of mind," MIT press, 1997

[6] Frith, Uta, and Francesca Happe, "Theory of Mind and Self-Consciousness: What Is It Like to Be Autistic?," Mind & Language 14.1 (1999): 82-89

[7] Fiannaca, A., Apostolopoulous, I., and Folmer, E. "Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces," Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility, pp. 19-26, 2014

[8] Hernandez, J., and Picard, R. W. "SenseGlass: using google glass to sense daily emotions," Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology, pp. 77-78, 2014

[9] Howse, J. "Android Application Programming with OpenCV," Packt Publishing Ltd

[10] http://www.mathworks.com/matlabcentral/fileexchange/44400-tutorial-and-toolbox-on-real-time-optical-flow

[11] https://developers.google.com/glass/develop/gdk/location-sensors

[12] Xu, Q., S.-C. Cheung and N. Soares. "Littlehelper: Using Google Glass to Assist Individuals with Autism in Job Interviews," in International Meeting for Autism Research (IMFAR), May 13-16, 2015. https://imfar.confex.com/imfar/2015/webprogram/Paper19529.html

[13] Ching, R. P. "Relationship between Head Mass and Circumference in Human Adults," Technical Brief, Applied Biomechanics Laboratory, University of Washington. Date: July 20, 2007.

[14] Viola, P. and Jones, M. J. "Robust Real-Time Face Detection", International Journal of Computer Vision 57(2), 137–154, 2004