A Wearable Social Interaction Aid for Children with Autism

Peter Washington
Stanford University
Stanford, CA 94305, USA
peterwashington@stanford.edu

Jena Daniels
Stanford University
Stanford, CA 94305, USA
danielsj@stanford.edu

Catalin Voss
Stanford University
Stanford, CA 94305, USA
catalin@stanford.edu

Carl Feinstein
Stanford University
Stanford, CA 94305, USA
carlf@stanford.edu

Nick Haber
Stanford University
Stanford, CA 94305, USA
nhaber@stanford.edu

Terry Winograd
Stanford University
Stanford, CA 94305, USA
winograd@cs.stanford.edu

Serena Tanaka
Stanford University
Stanford, CA 94305, USA
smtanaka@stanford.edu

Dennis Wall
Stanford University
Stanford, CA 94305, USA
dpwall@stanford.edu

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892282

Abstract
Over 1 million children under the age of 17 in the US have been identified with Autism Spectrum Disorder (ASD). These children struggle to recognize facial expressions, make eye contact, and engage in social interactions. Gaining these skills requires intensive behavioral interventions that are often expensive, difficult to access, and inconsistently administered. We have developed a system to automate facial expression recognition that runs on wearable glasses and delivers real-time social cues, with the goal of creating a behavioral aid for children with ASD that maximizes behavioral feedback while minimizing the distractions to the child. This paper describes the design of our system and interface decisions resulting from initial observations gathered during multiple preliminary trials.

Author Keywords
Autism; Behavioral Therapy; Wearable Computing; Ubiquitous Computing

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

Introduction

About 1 in 68 children have been identified with Autism Spectrum Disorder (ASD) [16]. These children struggle to recognize facial expressions, make eye contact, and engage in social interactions [5]. Gaining these skills requires intensive behavioral intervention that is often difficult to access and inconsistently administered. The current standard of care consists of "flashcard therapy" involving painstaking memorization of facial emotions [6]. As a result, many children with ASD fail to build core social skills and can quickly regress down a path of isolation that worsens their symptoms.

While computer-assisted in-place treatment systems have been studied for years [1,4,13], only a few strides have been made to bring the learning process away from flashcards and into the daily life of children with ASD. Previous attempts to design just-in-time in-situ learning aids for facial affect showed promising potential by presenting emotional cues in a fun and intuitive way on a mini-computer [9]. However, such systems did not achieve widespread use due to the large and perhaps obtrusive nature of the feedback aids and the state of expression recognition technology available at the time. The arrival of wearable devices such as Google Glass presents the opportunity to develop applications that give minimally obtrusive social cues in real time.

Our system

We have developed a system for automatic facial expression recognition that runs on wearable glasses and delivers real time social cues to individuals with ASD in their natural environment. With the goal of creating a behavioral aid that maximizes feedback while minimizing the distractions to the child, we built a prototype of our system on Google Glass. Our system tracks expressive events in faces using the outward-facing camera on the device. The system then gives social cues in real time, for example about facial expressions (i.e. "Happy," "Angry," etc.) shown on the heads-up display of the Glass. The system also can record social responses, such as the amount and type of eye contact using Glass' sensors and a custom-made infrared eye tracker.

Our system supports multiple tasks that can be integrated into a therapy program in various ways:

1. When the Glass is worn at home, our system is used as an unstructured emotional aid in social activities (e.g. dinner).
2. To provide more engaging ways of interaction, we include various gamified activities that the children can use during informal behavioral therapy sessions. During in-lab trials, we evaluated a game termed "Capture the Smile," adapted from Picard et al.'s work at MIT, in which a group of children is challenged to "capture" a certain number of selected emotions by provoking them in an adult who makes the face that contains the emotion, e.g., by making a compliment to provoke a happy face.
3. We have developed an Android application, which allows parents and children to review activities recorded throughout the day. "Emotional moments" are auto-curated and highlighted in videos. Parents are encouraged to review these moments with their children (and if they choose, potentially, behavioral therapists). Finally, statistical data extracted from video and sensory outputs are stored in this application and regularly transmitted to our research team. In the long run, families and therapists will be able to access these

Visual Feedback Cues

Text: The name of the emotion is written on the screen. The screen of the Google Glass contains a white background and the font color is black.

Color: Work in color psychology suggests that humans associate colors with certain emotions [14]. In color mode, the screen is filled with the color associated with the recognized emotion.

Smiley: A playful smiley representing the emotion is displayed on the screen.

Figure 1: Visual feedback cues.

reports (e.g. on eye contact insights) along with the video data.

This late-breaking work presents our system architecture and then discusses how observations from in-lab studies with 40 children between 6 and 17 years old (20 with ASD, 20 typically developing) as well as informal at-home trials resulted in design choices in developing our wearable aid. We propose multiple real-time feedback mechanisms for structured games and unstructured emotional feedback and provide a brief observational comparison. Finally, we give a preview of an upcoming formal 100-person at-home trial, which will evaluate the system in much greater detail.

System Architecture

Figure 3 provides an overview of our system. Each behavioral aid "unit" is comprised of a pair of wearable devices and an Android phone, which are connected via a secure wireless network. The wearable glasses act as a sensory input device for video and head pose data as well as an output device through a bone-conducting speaker, optional earpiece, and a heads-up display. To spare the limited battery life and processing power on the glasses, all computationally intensive tasks such as frame processing, video encoding, and data storage are handled on the phone. When users want to start or finish a new activity, they can use the Android app, which uses a custom protocol to trigger the right activity on the wearable glasses and prompts the Google Glass to begin capturing camera frames at a rate of approximately 30 frames per second. When a new frame is captured, the full frame and other sensor data as well as a high-resolution image of the face area are transmitted to the phone for processing. The phone encodes and stores the video data in a parallelized fashion and runs the frame through our expression recognition pipeline. It then transmits a result packet including metadata about the face(s) tracked and expressions to the glasses. The glasses interpret the result and show the appropriate social cue (either on the display or as audio). This process happens in real time with less than 50ms latency.

Emotion Recognition

Our real-time expression recognition system builds on a lightweight face tracker and a machine learning classifier that achieves a classification accuracy of 97% for 8 emotions on previously unseen subjects in the popular Extended Cohn-Kanade Dataset [8]. Our classifier has been trained on a collection of academically available datasets as well as some self-collected data and rendered 3D facial scans. It employs a "baseline correction method" [3] that we developed specifically for the flexible at-home setting and provides a practical performance gain across a wide variation of faces, head positions/poses, and lighting (e.g. from shaded regions to high light). Finally, we designed a method for person-specific calibration that uses a Bayesian update to adjust a general emotion recognition model to a specific family after quick training to make sure we achieve 97%+ accuracies and low false positive rates on family members with whom the child frequently interacts. The technical details of our expression recognition methods are available in [2,3,15].

Eye Tracking

To address the challenge of measuring eye contact, a skill with which children with ASD almost unanimously struggle, and to gather additional insights into the gaze pattern of children when using our device within the

Parent Review System

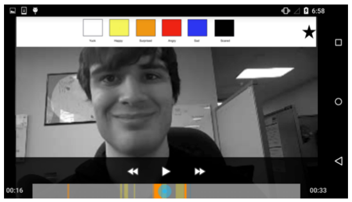

Figure 2: While reviewing the videos, parents have the ability to mark particular parts of the session as important. This can signify a variety of things to both the parents and researchers, such as an egregious error in emotion recognition, interesting behavior by the child in response to the behavioral therapy, or an otherwise important part of the video that should be reviewed at a later time.

laboratory, we designed and built a $35 infrared pupil-tracking camera that clips onto Glass based on the Pupil Labs Project [11]. This allows us to determine eye contact and mutual gaze events occurring between the wearer and a conversation partner during social interaction or analyze the gaze of participants at faces on a computer screen.

Feedback Mechanisms

There are several potential feedback mechanisms that can be used on the Glass. An ideal feedback mechanism will maximize behavioral information that the child receives while minimizing the distractions to various feedback mechanisms that we have designed and built.

Visual

We used three primary emotional visual feedback mechanisms: text, color, and an emoticon, as detailed in Figure 1. The visual feedback can be any combination of the 3 basic feedback mechanisms. However, the number of visual cues provided is kept to a minimum because children with ASD do not react well to sensory overload [10].

Audio

The audio feedback includes a narrator reading out the name of the emotion as well as a range of child-friendly playful sound effects associated with the emotion. In addition to the possibilities for visual and audio feedback individually, the cues can also be combined, allowing for a large range of feedback mechanisms.

Parent Review System

The parent review system is an Android application that runs on the same phone that performs the computationally intensive emotion recognition. As seen in Figure 2, the review system contains a newsfeed-like view of the previous session recordings in chronological order. The parents have the ability at any time to view, hide, or permanently delete videos from the newsfeed.

In-Lab Trials

In order to initially evaluate our system's feedback mechanisms, we recruited 20 children with ASD and 20 typically developing children and conducted experiments at the Stanford School of Medicine under approved Institutional Review Board (IRB) protocols and under the supervision of clinicians and the Director of the Stanford Autism Center.

The participants wore the Glass devices and went through a calibration step with qualified research staff. The calibration consisted of participants tracking a moving object with their eyes on a 30-inch computer screen for 30 seconds in order to optimize the eye tracking technology. The participants were then asked to view three separate batches of 7 standard facial emotions from the Child Affective Face Set (CAFE) dataset [7] balanced for varying facial expressions, genders, and races. In each batch, participants were asked to make a judgment on which emotion was being expressed on the screen, and his/her final response was recorded. The screen displayed the images for 6 seconds alongside two alternating social and non-social standardized distractor images from the "High Autism Interest" database [12]. In addition, we provided the participants with the list of 7 emotions (Happy, Angry, Sad, Disgusted, Afraid, Surprised and Calm) following every image.

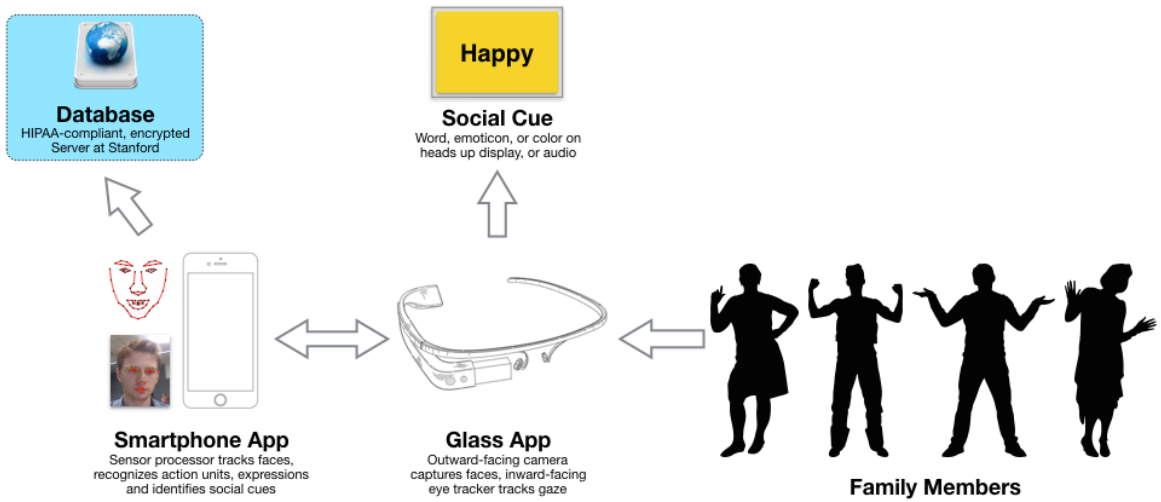

Figure 3: Overview of the Autism Glass system.

We provided the first and third batches to the participant without any emotion feedback. In the second batch, we enabled our social feedback system to deliver the correct emotion either in audio or visual form (as words displayed on the heads-up display on the Glass unit itself) after the face had been shown for 3 seconds. By doing this, we were able to track short-term learning opportunity the social feedback provided, as well as record any delta from the first and third batch. We made several observations during these trial sessions.

Children responded well to wearing the devices

Prior to the in-lab sessions, we had concerns that participants with ASD would find wearing the device to be uncomfortable and either refuse to wear it or otherwise not complete the study. Contrary to our initial beliefs, only 5 participants commented that the device was uncomfortable and still completed the study procedures. 65% of the participants commented that they found the Glasses to be comfortable. The biggest complaint about the device was that it became uncomfortably warm after being used for more than about 20 minutes. In order to resolve this, we will limit the amount of time that the Glass system is used both during future in-lab studies and in the upcoming at-home trials.

Children responded well to the images and enjoyed the gamified activities and feedback mechanisms

The more playful visual feedback cues (such as the emoticons) resulted in more children engaged with the

Batch	ASD	TD
1	85%	85%
2	94%	95%
3	88%	90%

Figure 4: Overall performance of correct emotion recognition from children with ASD (n=18) and typically developing (TD) children (n=20). The numbers exclude 2 children with ASD who did not complete all study procedures. A detailed analysis of these results will be presented in [2].

activity. The children with ASD who did not complete the study stopped not because of the device but rather due to the monotony of the in-lab task.

Children tend to respond better to audio feedback than to visual feedback
We used only visual cues for our first 6 participants (typically developing). They commented that it was distracting and difficult to view. For all other participants (both ASD and typically developing), we used only audio cues. 100% of the participants who had the audio cues commented that it was clear, easy to hear, and not a distraction.

The majority of children went with audio cues over their own intuition
Of the 30 participants who were asked, 11 children with ASD and 10 typically developing children said they went with the social cues from the Glass over their own intuition. In addition, 5 children with ASD and 2 typically developing children went with their intuition over the Glass cues.

There is some indication that there may be differences between the children with ASD and typically developing children
Figure 4 shows the overall performance of emotion recognition from children with ASD and typically developing children. A detailed analysis of these results will be presented in [2].

Future Work
The initial in-lab trials revealed that our system could be used as part of an effective behavioral therapy for children with ASD. However, the greatest benefit of the system will come from prolonged at-home use. We have only tested the system at home with 2 typically developing children and have yet to test the system at home with children with ASD. We have recruited 80 children with ASD and 20 typically developing children to partake in a 4-month IRB-approved at-home trial.

During the trial, we will explore alternative feedback mechanisms in an unstructured setting. The ideal feedback mechanisms for an unstructured activity at home could be vastly different from those in the lab setting. For example, while a child may perceive the audio feedback as fun during an in-lab game, repeated audio feedback is likely to be extremely distracting in a conversational setting at the dinner table. We also want to test the parent review system as a potential behavioral therapy mechanism. Parents will be encouraged to review the videos with their children. The color-coded video seek bar provides additional emotional feedback to the child while watching the video and reinforces the color-to-emotion mappings.

Acknowledgements
This work was supported with funding from the Hartwell Foundation, Dekeyser and Friends Foundation, and by the David and Lucile Packard Foundation Special Projects Grant under grant 2015-62349. We would also like to thank Google for donating 35 units of Google Glass through the Google Gifts Material Grant.

References
1. Bosseler, Alexis, and Dominic W. Massaro. "Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism." *Journal of autism and developmental disorders* 33.6 (2003): 653-672.

2. Haber, Nick, Catalin Voss, Azar Fazel, Jena Daniels, Serena Tanaka, Terry Winograd, Carl Feinstein, Dennis Wall. "Autism Glass: Wearable Behavioral Aid Augments Social Learning in Children with Autism" in submission, 2016.
3. Haber, Nick, Catalin Voss, Dennis Wall, Azar Fazel, and Terry Winograd. "A practical approach to real-time neutral subtraction for facial expression recognition" accepted in: IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.
4. Hetzroni, Orit E., and Juman Tannous. "Effects of a computer-based intervention program on the communicative functions of children with autism." *Journal of autism and developmental disorders* 34.2 (2004): 95-113.
5. Jones, Warren, and Ami Klin. "Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism." *Nature* 504.7480 (2013): 427-431.
6. Lerman, Dorothea C., et al. "A rapid assessment of skills in young children with autism." *Journal of Applied Behavior Analysis* 37.1 (2004): 11.
7. LoBue, Vanessa, and Cat Thrasher. "The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults." *Frontiers in psychology* 5 (2014).
8. Lucey, Patrick, et al. "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression."*Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010.
9. Madsen, Miriam, et al. "Technology for just-in-time in-situ learning of facial affect for persons diagnosed with an autism spectrum disorder." *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2008.
10. Myles, Brenda Smith, et al. "Sensory issues in children with Asperger syndrome and autism." *Education and Training in Developmental Disabilities*(2004): 283-290.
11. PupilLabs, https://pupil-labs.com/pupil/.
12. Sasson, Noah J., and Jed T. Elison. "Eye tracking young children with autism." *Journal of visualized experiments: JoVE* 61 (2012).
13. Silver, Miriam, and Peter Oakes. "Evaluation of a new computer intervention to teach people with autism or Asperger syndrome to recognize and predict emotions in others." *Autism* 5.3 (2001): 299-316.
14. Valdez, Patricia, and Albert Mehrabian. "Effects of color on emotions."*Journal of Experimental Psychology: General* 123.4 (1994): 394.
15. Voss, Catalin, Nick Haber, Dennis Wall, and Terry Winograd. "Automated discovery of facial action units using deep convolutional neural networks" in submission, 2016.
16. Wingate, Martha, et al. "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010." MMWR Surveillance Summaries 63.2 (2014).