SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism

PETER WASHINGTON, Stanford University
CATALIN VOSS, Stanford University
AARON KLINE, Stanford University
NICK HABER, Stanford University
JENA DANIELS, Stanford University
AZAR FAZEL, Stanford University
TITAS DE, Stanford University
CARL FEINSTEIN, Stanford University
TERRY WINOGRAD, Stanford University
DENNIS WALL, Stanford University

We have developed a system for automatic facial expression recognition running on Google Glass, delivering real-time social cues to children with Autism Spectrum Disorder (ASD). The system includes multiple mechanisms to engage children and their parents, who administer this technology within the home. We completed an at-home design trial with 14 families that used the learning aid over a 3-month period. We found that children with ASD generally respond well to wearing the system at home and opt for the most expressive feedback choice. We further evaluated app usage, facial engagement, and model accuracy. We found that the device can act as a powerful training aid when used periodically in the home, that interactive video content from wearable therapy sessions should be augmented with sufficient context about the content to produce long-term engagement, and that the design of wearable systems for children with ASD should be heavily dependent on the functioning level of the child. We contribute general design implications for developing wearable aids used by children with ASD and other behavioral disorders as well as their parents during at-home parent-administered therapy sessions.

1 INTRODUCTION

Autism Spectrum Disorder (ASD) is quickly becoming a global health crisis, estimated to affect 1 in 68 children [5] and cost the U.S. health system over $137 billion each year [29]. The neurodevelopmental disorder results in

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 1, No. 3, Article 112. Publication date: September 2017.

11

impaired communication, eye contact, expression recognition and social interactions [10, 11, 12, 27, 28, 32, 33, 41, 48, 60]. Today's best-known intervention, applied behavioral analysis (ABA), relies on teaching these skills in clinical environments, largely removed from where they will be used and relying on tools such as flashcards [33]. While this intervention can lead to improvements [34], the therapy is applied inconsistently and often generalizes poorly to situations that go beyond the routines practiced with the relatively limited clinical tools available [10].

Even when programs are properly administered, their delivery is increasingly bottlenecked as the number of available therapists lags well behind by the number of children in need of care. Wearable aids such as Google Glass provide a means of helping children with ASD because they allow for the juxtaposition of the user's field-of-view with virtual visual and audio feedback, which creates an opportunity to reinforce real-world concepts such as a person's emotional state while having a conversation with another person. We have developed an artificial intelligence tool for automatic facial expression recognition, named SuperpowerGlass, that runs on Google Glass and delivers instantaneous social cues to children with ASD *in their natural environment*. We tested the system via a longitudinal at-home study under the supervision of the Stanford Institutional Review Board. All identifiable data we collected are secured in HIPAA compliant systems.

At a high level, the computer vision system uses the glasses' outward facing camera to read a person's facial expressions by passing video data to an Android native app for immediate machine learning-based classification of the commonly used Ekman "basic" emotions for which training-data are readily available: anger, disgust, fear, happiness, sadness, and surprise, plus a contempt (or "meh") and a neutral category [13]. The system then gives the child wearer real-time social cues and records social responses, including the amount of eye contact and level of social engagement. Through a dedicated app, parents can then review and discuss videos of social interaction. These videos are auto-curated to enable fast navigation by highlighting the emotions detected in the video on the video playback bar.

Rather than acting as a prosthetic that is constantly worn by the child, the system is meant to be used as a training aid that is used for a few minutes to an hour on a regular basis. An important aspect of this system is that the experience must be engaging not only for the child with ASD but also for the parent, who must consistently and repeatedly use the wearable system with their child. This introduces a vast array of additional design constraints, as all design decisions must be considered in terms of both the child and the parent.

To evaluate the effectiveness of our wearable system and to determine how families use the system over a twelve-week period at home, we launched an iterative design study with 14 families with a child impacted by ASD. These families took the devices home and used them on a regular basis. Although the families were asked to use the device for an hour every day, there was variation in the participation rate of the families. The study produced over 5,000 minutes of app usage data from the 14 families, which we used to learn about the effectiveness of various parts of the system. We tracked emotional learning progress with periodic in-lab check-ins while continuously gathering device data. Along with observations made by parents, the data suggest that the therapy delivered by our system has the potential to facilitate substantial improvement in children's social skills. This methodology enabled us to refine design choices about the system while learning about each family's experiences with the system.

Our contributions to HCI are (1) a wearable emotion-recognition and feedback system for children with ASD that is designed for regular periodic administration by parents in an at-home setting and (2) numerous design lessons learned from a longitudinal at-home iterative design trial involving families with children with varying severities of ASD.

2 RELATED WORK

2.1 Interactive Assistive Technology For ASD

Head-worn displays have been used to provide unobtrusive, always-available, and glanceable vocabulary support to individuals with Aphasia. It was found that a head-worn approach better allows wearers to maintain

focus on the conversation, reduce reliance on the availability of external tools (e.g., paper and pen) or people, and minimize visibility of the support by others [59].

Interactive visual displays have been developed for children with ASD. vSked, an interactive and collaborative visual scheduling system designed for children with ASD in elementary school classrooms, was found to reduce staff effort for using the visual cues and also resulted in improvements in the perceived quality and quantity of communication and social interactions in the classroom [26, 61].

Techniques that involve verbal and visual prompting annotated on top of the physical objects used during therapies for children with ASD have also been explored in Mobis, a mobile augmented reality application that lets teachers superimpose digital content on top of physical objects [16]. MOSOCO, another mobile assistive technology, uses the visual supports of a validated curriculum to help children with ASD practice social skills in real-life situations [15].

Hayes et al. have determined design requirements for assistive technologies that use interactive visual supports to engage students and support parents simultaneously: flexibility, communication and collaboration capabilities for both children and parents, and parent support for programming and documentation of use [24].

It should be noted that although visual support systems have been extensively explored for children with ASD, audio coaching has been shown to be an effective instructional tool for children with ASD as well [4]. We therefore explore both visual and audio cues in SuperpowerGlass.

2.2   Wearable Technologies For ASD

While there are numerous concerns regarding user perceptions of smart glasses for general-purpose use [31], other studies have found that participants consider head-mounted display use more socially acceptable when the device is being used to support a person with a disability [46]. Because of the social acceptability of wearable displays for disabilities and the large potential for wearable technologies to aid those with disabilities, there is a great deal of literature on wearable systems to aid children with ASD.

Virtual reality systems using a fully immersive headset have been found to be effective learning aids for both children and young adults [29]. Parent-driven use of wearable cameras for ASD support has shown that the ability to see the world through their child's eyes was of high value to parents. The camera supported parents in understanding their children's needs and helping their children engage with the world [37]. Hayes et al. have found that automated capture can be successfully applied in a variety of settings to assist with the education and giving of care to children with ASD, while also providing a means of keeping records of those activities [23].

Picard's lab at MIT has explored the use of wearable systems and affective computing as companion tools for social-emotional learning and the use of the recorded videos for defining a process to collect, segment, label, and use video clips from everyday conversations [44, 55]. The group has built a prototype of a Google Glass application to sense daily emotions in users, identifying several challenges of using Google Glass to detect emotions, such as constrained battery life, limited storage space, and privacy concerns over transmitting affective information over a wireless network [25].

Various wearable systems have been developed specifically for autism support. Boyd et al. have developed SayWAT, a wearable assistive system which can automatically detect atypical prosody and deliver feedback in real time without disrupting the wearer or the conversation partner [7]. They found that for short-term engagement, visual wearable feedback is likely to be appropriate for people with ASD, but that for longer-term use, additional work is needed to understand the potential opportunities and challenges of visual wearable feedback [7]. ProCom has been developed as a wearable assistive technology which can support people with ASD in developing awareness of physical proximity in social settings; the researchers found that children with autism respond to real-time information in wearable visualizations during a conversation [8]. We use this insight in the development of our system which provides emotion feedback to children during a conversation or other person-to-person interaction.

None of these wearable systems for children with ASD are meant to be used for a prolonged period in an at-home setting while regularly being administered by parents. We extend this prior work on wearable aids for children with autism by exploring challenges that arise when the system is frequently used at home over a prolonged time period and controlled by the parent. To address this design challenge, we introduce several customizable options for visual and audio feedback to promote continued parent and child engagement, a facial engagement indicator to promote face-to-face interaction during the sessions, and a session review application that highlights emotional moments during video playback of a previous therapy session. These additions have allowed us to make several design insights about wearable systems for children with ASD used at home and administered by parents.

2.3 Behavioral Therapy Systems for ASD

Recent work explores technology to enhance therapy for children with ASD. The ECHOES project explores a technology-enhanced learning (TEL) environment that facilitates acquisition and exploration of social skills by typically developing children and children with ASD [45]. Another study found that watching a television series designed to enhance emotion comprehension every day can improve the emotion comprehension of young children with ASD [21]. Multitouch interfaces for behavioral therapy have been found to increase collaboration skills in children with ASD [3, 19, 20]. Humanoid robots used as catalysts for social behavior have been shown to promote positive child-human interactions [17]. Much of this work has determined that children with ASD enjoy gamified systems and that a playful system is generally most effective in providing useful behavioral therapy. Frameworks for designing games to teach emotion to children with ASD have also been successful [43]. Robot therapy for ASD has been explored as one of the first applications for socially assistive robotics. Scassellati et al. (2012) found that social robots may be developed into useful tools for social skills and communication therapies, specifically by embedding social interaction into intrinsic reinforcers and motivators for children with ASD [50]. Ahmed et al. have found that for children with autism, behavioral engagement (the proportion of time a participant has their face oriented to the computer screen) and emotional engagement (the activation of facial activation units) are strongly correlated with one another, but that emotional (not behavioral) engagement is a better predictor of test performance [1]. Our system takes a different approach to addressing this phenomenon by providing feedback reinforcing behavioral engagement for the child only when the child elicits emotional engagement with a conversational partner.

Kientz's work has determined several relevant principles for designing ubiquitous systems for children with ASD: changes in behavior should be seamless for the child, easier systems are better, customization of the system is critical, and parents are usually a better source of feedback about the system than the children themselves [23]. Participatory evaluation with children with ASD has been explored and found to be successful through dedicated evaluation phases including the co-definition of goals and methods, joint processes of data gathering and the co-interpretation of results [52]. Spiel et al. have developed an approach for capturing and describing the multi-faceted experiences autistic children have with technologies, drawing on diverse data sources ranging from interviews to log-data, and the first-hand perspective of autistic children [53]. We used these various data sources in our evaluation of SuperpowerGlass in order to understand the multi-faceted experiences that the system may result in at home as opposed to in lab.

We extend the study of behavioral therapy systems for ASD by conducting, to our knowledge, the first longitudinal study observing *prolonged use* of a wearable device in an *at-home setting* by children with ASD. We identify several differences required in the design of such therapy systems that arise when the devices are used at home and controlled by parents (as opposed to in-lab studies).

3 SUPERPOWER GLASS SYSTEM DESIGN

3.1 Overall Architecture

The overall system architecture for the SuperpowerGlass system is depicted in Figure 1. The system consists of two components: the Google Glass (worn by the child) and a wirelessly linked Android phone (used by the parent). The system is controlled by the Android application, which allows the parent to initiate one of several activities (detailed in the next section). During a session, social cues are delivered to the child in real-time on Glass based on the results of the emotion recognition model that runs on the Android device. A video recording of the session is saved on the Android device. During the initial pre-trial interviews, parents explained that they were comfortable with sharing video recordings of at-home recordings but were not willing to share the audio of their conversations. Due to these initial privacy concerns, the audio from the activity sessions is not recorded. After a session is complete, the parent and child can review a video recording of previous sessions on the Android app.
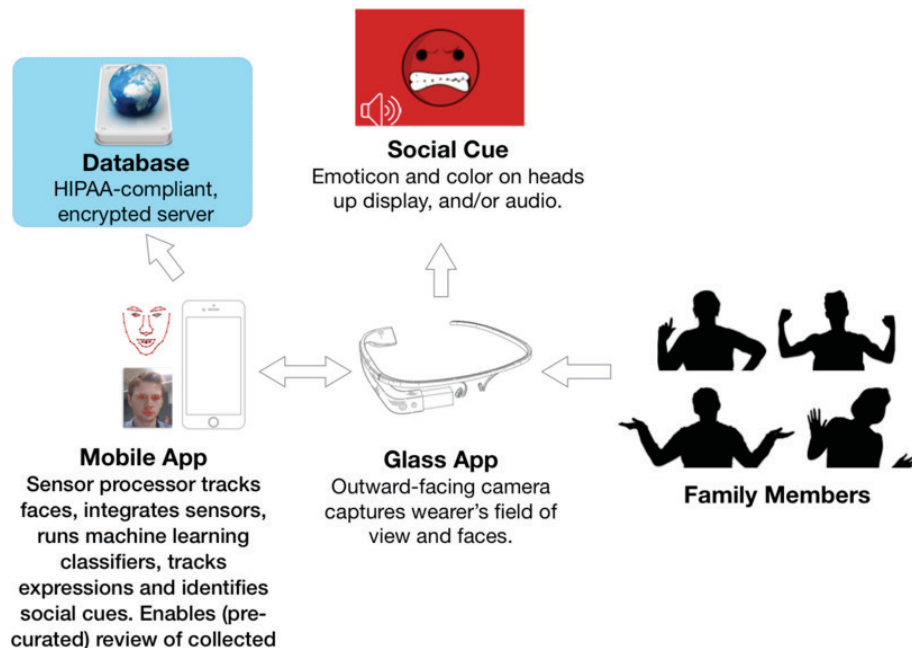

Fig. 1. Overview of the SuperpowerGlass system.

Fig. 2. The home screen allows parents to scroll through a newsfeed of previously recorded sessions or to start a new session. Parents can select various parameters about the session, such as the exact activity, the type of feedback (audio or visual or both), and the type of facial engagement indicator returned to the heads-up display in real-time.

3.2 Activities

The Android application allows the user to select between three different types of activities: two structured game-like activities (Capture the Smile and Guess the Emotion) and an unstructured activity where a social cue is provided whenever the corresponding facial expression is detected on the Glass. Multiple games are added in order to increase engagement by increasing the selection of activity choices. We also wanted to determine whether children with ASD will prefer cycling through a series of games over a long period of time or converge to a preferred game.

3.2.1 Capture the Smile. The first structured activity, "Capture the Smile", is a game adapted from Picard's work at the MIT Media Lab [14] in which the child is challenged to provoke prompted emotions in an adult who makes the face corresponding to the emotion. For example, the child can give a compliment to the adult in order to provoke a happy face. The glasses play an audio prompt of the form "find a/an <emotion> face". Once the emotion recognition system recognizes the emotion on the adult's face, audio or visual feedback is provided to the child via the glasses (the feedback mechanisms are described in more detail in the next section). Our implementation of Capture the Smile lasts for 180 seconds or until the child has successfully evoked 20 emotions, whichever comes first. If the adult and child are stuck and cannot evoke an emotion, they have the option to swipe forward on Google Glass' touchpad located near the wearer's right temple.

3.2.2 Guess the Emotion. The second structured activity, called "Guess the Emotion", challenges the child to guess the emotion evoked on the face of an adult. In each round, the adult initially selects on the Android app the emotion they want to evoke. After the adult has chosen the emotion and made the corresponding face, the child verbally guesses the emotion. If the child guesses the correct emotion, then the glasses play/display the social cue corresponding to the emotion. The game is scored but can last for as many rounds as the adult and child wish to play the game.

3.2.3 Unstructured Activity. When the glasses are worn at home, they can be used as an unstructured emotional aid during social activities (e.g., dinner). During the unstructured activity, the glasses provide the selected feedback mechanism whenever it recognizes the emotion. The user simply wears the Google Glass system, receiving feedback whenever the system detects an emotion (rather than in the context of a game like Capture the Smile of Guess the Emotion). To prevent excessive feedback, the glasses provide the social cues at a maximum rate of once per six seconds for repeated emotions. The unstructured activity has no time limit. Families are provided a list of suggested activities divided into three categories ("Play", "Tell me...", "Others"), including "Free Play" and "Other" options from which they choose an action. Their choice has no impact on the way the system works during the activity, but is recorded for analysis.

3.3 Feedback Choices

Several potential feedback mechanisms can be run on the glasses, in the form of both visual and audio cues. An ideal feedback mechanism will maximize behavioral information given to the child while minimizing distraction to the child's focus. This is a particular concern with children with ASD, who are believed to hyper-focus on nonsocial objects and react negatively to sensory overload [40]. We therefore have explored various visual feedback mechanisms to display on the glasses, each with varying levels of potential distraction to the child. The visual feedback can also be combined with audio feedback for even more stimulation. Before a session begins, the parent has the option to select the feedback mechanism that will be used during the session. Our aid allows users to choose between three emotional feedback mechanisms: visual cues, auditory cues, and combined visual and auditory cues.

3.3.1 Visual Emotion and Color Cues. For the visual feedback, [57] introduces three possible heads-up interfaces and describes in-lab interaction observations that argue for a combined emoticon and color feedback as depicted in Figure 3. This can be justified with perception literature. It has been shown that a majority of children with ASD are able to recognize the emotion displayed by a digital avatar [39]. Work in color research suggests that people notice colors before shape or text within their peripheral field of view and tend to associate colors with certain emotions [56]. While not all of the emotion-to-color associations are universal [30], children tend to make similar associations between colors and emotional concepts [6]. We selected the colors for each emotion using the guidelines presented in [6]. Because we were only studying 7 emotions, we hypothesized that repeatedly showing a color with an emoticon and auditory feedback would allow children to eventually notice the social cues within their peripheral field of view, rather than having to actively look at the screen, and thus provide an unobtrusive cue.

Fig. 3. There are various potential visual feedback cues that can be displayed to the user.

3.3.2 Auditory Cues. Because audio coaching has been shown to be an effective instructional tool for children with ASD [4], we decided to provide users the option before beginning a session to receive audio feedback either in addition to or instead of visual feedback. The audio feedback consists of a female voice reading out the name of the emotion. The audio feedback is emitted via a bone-conducting speaker on the glasses, which is loud enough for the child to hear and only faintly noticeable by a conversation partner in a home environment. Children also had the option to use a right-sided mono earpiece that plugs into the glasses for improved sound quality and privacy, but mostly did not find this necessary.

3.4 Face-Tracking Indicator

During the in-lab trials conducted in [58], situations were observed in which the system would fail to recognize emotions not because of a problem with the recognition model but because a face was not detected, for example due to challenging backlighting. To enable users to distinguish between a lack of cues and a correctly tracked but neutral face, we have added an indicator signal on the display of the glasses that lights up whenever a face is detected. We explored whether the presence of the indicator would result in increased facial engagement over time with periodic use of the SuperpowerGlass system. The indicator is bright green, primarily because green is not one of the colors used to represent any of the emotions. As shown in Figure 4, there are four indicator options. Each indicator option varies in its level of prominence (area of screen coverage). The indicator is

generally visible to both the child and the parent due to the clear display of Glass. The user is prompted to select one of the four choices at the beginning of each session. In order to see whether child preferences for indicator design stabilizes over time, we asked the parents to allow the children to choose the indicator option which they preferred. For lower functioning kids who could not respond/not choose the indicator option they liked the most, the parents chose for them.
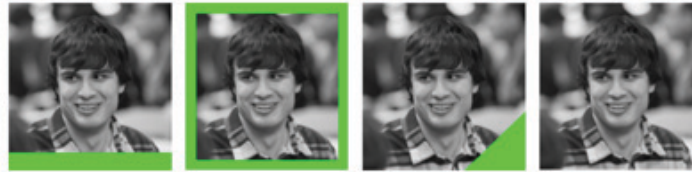

Fig. 4. Face-tracking indicator options.

3.5 Parent Review System

In order to facilitate continued parent and child engagement with SuperpowerGlass, we have added a parent review system on the Android app which allows parents and children to review previously recorded sessions. The review system contains a newsfeed-like view of the previous session recordings, as shown in Figure 5. In the video playback for each video, emotions are gathered from metadata that are stored with each frame of the recorded video so that the emotions displayed on the video playback bar are aligned with the emotions that were captured during recording-time, as shown in Figure 6. This allows users to skip to relevant parts of the video. Parents are encouraged to review videos of past sessions with their children. While reviewing the videos, parents can mark particular parts of the session as important. This serves as a bookmarking tool for parents and researchers. The parent review feature capitalizes on the longitudinal nature of the system usage, allowing parents to point out to children differences in behaviors over time.
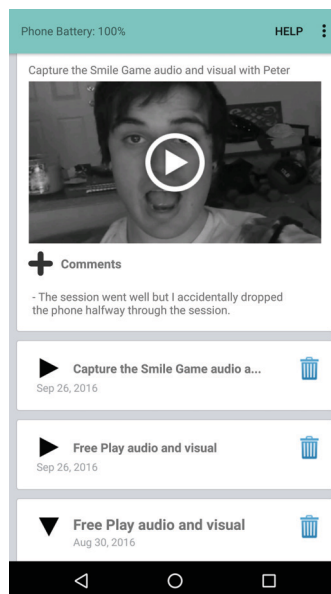

Fig. 5. The parent review application contains a newsfeed-like view of the system.
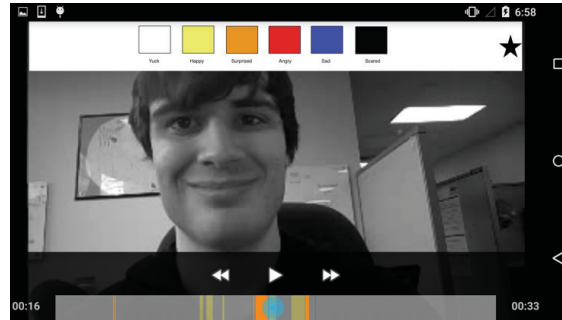
Fig. 6. While reviewing the videos, parents can mark parts of the session as important by clicking the star button on the upper right-hand corner of the screen. This can signify a variety of features important to either/both the parents and researchers, such as an error in emotion recognition, notable behavior in the child's response to therapy, or an otherwise important part of the video that should be reviewed at a later time.

4 IMPLEMENTATION DETAILS

4.1 Communications Infrastructure

The glasses and Android device are connected via a dedicated wireless network, originating from the Android device and secured via WPA2 encryption. Users initiate a new activity via the Android application. This prompts the glasses to begin the activity and start capturing video frames. To reduce the video latency, the glasses send the uncompressed video frames to the phone via UDP as uncompressed greyscale frames at 320x194 resolution with a frame rate of 30 fps. While some frames may be dropped due to the nature of the UDP protocol, this is not of concern, as detected emotions are associated with a particular frame rather than a time.

During a session, the Android device displays a real-time preview of the incoming video feed and runs each frame through our expression recognition pipeline (detailed below). It then returns the detected emotion and face location data via UDP to the glasses. The glasses interpret the result and show the appropriate social cue (either on the heads-up display or as audio).

4.2 Facial Expression Recognition

Our real-time expression recognition system running on the Android device builds on a lightweight face tracker and an emotion recognition classifier. The system is built to be modular, so that the facial expression recognition classifiers are trained separately from other components of the computer vision system. This allows us to not only quickly evaluate the experience had with different models, but also adapt the system to different face-relevant tasks.

Our emotion recognition system consists of a logistic regression classifier that evaluates HOG features [9] of a face registered via a lightweight face tracker and made lighting-robust by the method of Tan-Triggs [54]. The system uses *neutral subtraction* [22], a method that learns the subject's neutral face at runtime in order to better discriminate expressions. After classification, we temporally filter the results in order to reduce the fire-rate of social cues and avoid false positives or rapid switching between cues if a classifier finds the data near the decision boundary. As we know who will be interacting with the user of the device for this use case, we optimize each participant's model to these individuals, adapting from a base model. The base model is trained on a mix of the Binghamton University 4D Facial Expression Database [62], the Bosphorus Database [49], the CMU Pose, Illumination, and Expression Database [51], the Extended Cohn-Kanade Dataset [35], the Japanese Female Facial Expression Database [36], and the MMI Facial Expression Database [42], along with our own data gathered in-

lab via another approved protocol through Stanford University Institutional Review Board. At training time, data are augmented by mirroring images and adding Gaussian field noise to the face tracker to simulate tracking inaccuracy.

User-specific models are then employed via hierarchical Bayesian domain adaptation [18], putting a Gaussian prior on the distribution of model coefficients. Adaptation can be done online via an approximation. Data are collected through an iterative training technique in which we ask subjects to act out various expressions multiple times, with acting-out queries made in an active fashion – the current model decides what it needs to see more of. This process usually takes 2-3 steps. Data are loosely curated, keeping all that the subject decides is a reasonable representation of the emotion, so long as there is a noticeable distinction made between expressions. An evaluation of this adaptation system is presented in the section on Model Accuracy.

4.3 Increasing Facial Resolution

Given the limited processing power and battery life of Glass, all computationally intensive tasks such as frame processing, video encoding, and data storage are handled on the Android phone. Despite this, the combination of camera processing and network activity on Glass still results in a limitation on the resolution of images that can be processed and transmitted at a frame rate sufficient for the emotion recognition processor. Further, this limitation increases over time, as the device temperature increases and Glass enacts CPU throttling to avoid overheating. To provide a sufficient frame rate through the entire life span of a real-time interaction session, but still increase the resolution available for emotion recognition, a dual approach is taken. Instead of providing the same image source for all conditions, a lower resolution full-frame image is provided whenever a face has not been detected; this allows for face detection through the camera's full field of view. However, when a face has been detected, the face's location in the field of view is used to instead provide a cropped view of the face from a higher resolution image. This allows for providing increased facial resolution to the emotion recognition processor while still working within the same processing and transmission limitations encountered on Glass.

4.4 Moving Window Frame Buffer

To avoid displaying the emotion cue for temporary errors in the emotion recognition model, the expression recognition server running on the Android employs a series of moving average filters during various stages of processing, including in the final expression classification, before sending a result back to the Glass device.

5 SYSTEM EVALUATION

5.1 Study Methodology

We conducted a 3-month long at-home iterative design trial exclusively with children with varying degrees of ASD and their parents in order to learn about the effectiveness of various aspects of the system, such as the emotion recognition model, the activity types, the feedback mechanism options, and indicator design. The average age in years for the 14 participants was 9.57 (SD=3.37, min=4, max=15), with n=3 (21.42%) females and n=11 (79.58%) males. Out of our 14 participants, 6 were Caucasian (including 1 Hispanic/Latino participant), 7 were Asian, and 1 was of Pacific Island origin. We asked each family to take the system home and use it for at least 20 minutes 3 times per week. Beyond this, they were encouraged to use it as they saw fit for any additional length they preferred.

The study was under the supervision of the Stanford Institutional Review Board (protocol 34059). All identifiable data we collected are secured in HIPAA compliant systems. Every participant went through an initial onboarding session in lab. During these onboarding sessions, families were given an overview of the system, including a trial-run of Capture the Smile, Guess the Emotion, and the Unstructured Activity. Facial emotion data were then collected from the parents and nuclear family members, with our baseline model being iteratively trained until convergence for all basic Ekman emotions. To provide an additional layer of security to the recorded videos on the Android application, all devices were locked with a 4-digit passcode provided to the

parents during this onboarding session. To correlate data about the participants' system usage with the functioning level of the child, each selected participant took the Abbreviated Battery Intelligence Quotient (ABIQ), which measures the areas of nonverbal fluid reasoning and verbal knowledge and includes two of the most important abilities predictive of academic and vocational advancement [47]. We interviewed each participant and their parents during these onboarding sessions, where we were able to make initial design modifications to the system before sending it home with the family.

Roughly every two weeks, the families returned to our lab for a check-in. During each check-in, we provided parents with an updated version of the Google Glass and Android apps. If the parents noticed a problem with the emotion recognition model, we would update the model until it consistently recognized all emotions correctly.

During these scheduled check-ins, a clinical examiner interviewed both the parents and child with ASD. The same clinical examiner conducted all of the interviews. Interviews lasted between 30-45 minutes and included a parent-directed and a child-directed set of yes-or-no questions (see Appendix A for the full list of questions). A clinical autism expert pre-approved the child-directed questions. The parent(s) were interviewed first, followed by the child. After each question was asked, the interviewee had the opportunity to elaborate on their response or to move on to the next question. If the interviewee chose to elaborate on their response, the clinical examiner noted verbatim quotes in a log containing all of the participant quotes. While the optional participant quotes were recorded during these intermediate check-ins, the responses to the yes-or-no questions were not recorded until the final conclusion appointment. Two kids were nonverbal during conclusion interviews. In both cases, the kid-directed set of questions were redirected to the parent, who would answer on behalf of the child.

We asked the same set of questions to each set of families during the final conclusion appointment, as well as an additional set of questions (see Appendix B for the full list of additional questions). For the additional set of questions, formal tallies were recorded and used for post-study synthesis by the study researchers. Parents and children were kept in the same room for both the check-in and conclusion interviews. As a result, the parents and children would sometimes elaborate upon each other's responses.

The quotes and qualitative findings presented in this paper are drawn from the interview content. Given the nature of this pilot design study with a small sample, we followed an iterative, inductive approach to finding patterns that pointed to potentially relevant outcomes and design issues. Members of our team reviewed the system logs and identified areas where there appeared to be meaningful patterns to explore further. Through the interview process described above, as well as a review of session videos and analysis of logged study data (detailed in Section 5.2), we made several observations about the use of the system outside of the lab, giving us several insights about the effectiveness and usability of wearable social aids outside of the lab. We reviewed (watched as a team) individual videos to illustrate points from the parent interviews as they were brought up, usually amounting to reviewing approximately 20 minutes of video data per participant. When parents marked videos as important via the video review system, we selected these videos to watch. If not, we asked parents if there were any notable videos to review. If parents did not have an answer, we selected videos with high levels of emotional content (which our system automatically logged, as described in Section 3.5) to review.

Table 1. Participant demographic data and scores on various screening questionnaires.

ID	Age (years)	Gender	Race	Diagnoses	SCQ Score
AG01	10	Male	White	ASD	26
AG02	10	Male	White	ASD	14
AG03	15	Female	White	ASD, PDD-NOS, ADHD, anxiety, depression, learning disability	29
AG11	4	Female	Native Hawaiian/ Pacific Islander	ASD	16
AG13	15	Male	Asian	ASD	17
AG14	7	Male	Asian	ASD	23
AG15	12	Male	White	Asperger's/ADHD	15
AG16	7	Male	Asian	Asperger's	29
AG17	11	Male	White	ADHD/Dysgraphia	15
AG18	11	Female	White	ASD/ADHD	31
AG29	10	Male	Asian	ASD	12
AG30	8	Male	White	ASD	23
AG31	10	Male	Hispanic/ Latino	ASD	31
AG37	4	Male	Asian	ASD	22

5.2 System Usage

To measure usage across interface choices (different indicators, activities, and feedback options) we encouraged participants to try out all options at home. During the biweekly check-ins, if a participant stated that they did not use a particular feature or only used a particular feature once, we asked them to use it during the next two weeks so that we could gather more concrete data about their interactions. We collected all recorded video data and marked sections of the videos along with custom usage logs during check-in days. These usage logs were recorded automatically by the SuperpowerGlass system and stored locally on the Android device until the check-ins, when the data was migrated to a HIPAA-compliant server using a custom analysis tool suite implemented in Python.

In total, we collected 5726 minutes of app usage log data. Participant mean usage of the app was 409 minutes (SD=266.8), which includes time spent with features like the parent review. Participants produced a total of 3017 minutes of session video data, with a mean amount of 216 minutes per participant (SD=162). The mean video session length was 5.28 minutes (SD=6.15).
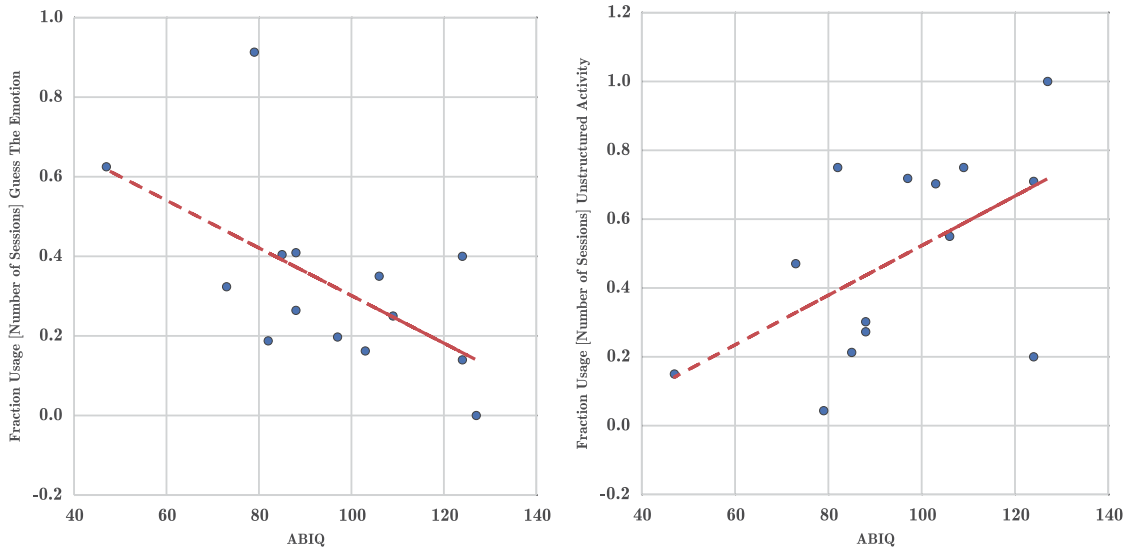
Fig. 7. Fraction of participant's usage that was Guess the Emotion and Unstructured Activity vs. functioning level (as measured by ABIQ) of each participant. The plot for Capture the Smile is not provided since Capture the Smile was a fixed-length game (whereas the user can play Guess the Emotion and the Unstructured Activity for as long as they wish).

While overall, families chose evenly between structured and unstructured activities, we noticed that this choice differed across the functioning spectrum of our participants. Figure 7 shows breakdowns of the activity choice for each participant along with their functioning level, revealing a downward trend in usage of the Guess the Emotion activity and an upward trend in usage of the Unstructured Activity with the increasing functioning level of the child, as determined by the ABIQ scale. Multiple families of lower-functioning children remarked that the structured games elicited more response from their child. Parents believed this was due to the fact that the games had structured goals to achieve, and clear predictable feedback responses when milestones were met. Families of higher-functioning children remarked that their child gravitated to the unstructured activity. One such family remarked that when their child was asked to try Guess the Emotion (one of the structured games), they tried it once, figured out how to play it immediately, accumulated a very high score, and then declared that the game wasn't challenging and that they had finished it. Another commented that their child didn't find the structured games challenging, and wanted to use the unstructured activity so they could choose their own game to play.

5.3 Facial Engagement

To quantify social interaction across different usage and activities, we employed a face-tracking based engagement measure we term "facial engagement". Due to the obtrusive nature of wearable glass based eye trackers described in [61], we did not utilize a full field-of-view eye tracker in this at-home study and instead relied on face tracking data extracted from the front-facing camera of Google Glass to identify video frames in which the child was engaging with a face. Specifically, we report normalized facial engagement for a given session video as:

$$\frac{\text{Frames containing face with yaw angle between}}{\text{-30 and 30 degrees}}$$
$$\frac{}{\text{Total number of frames}}$$

We hypothesized that this is a reasonable proxy measure for social overtures and perhaps even eye contact. We plan to further validate this using structured lab observations. Engagement measured this way exhibits relatively large standard deviations across sessions due to vastly heterogeneous usage patterns and the noise associated with face tracking in unstructured, naturalistic videos.

Although generally parents and children chose relatively evenly between structured and unstructured activities, facial attention was not even across both types of activities. Structured interactive games yielded more facial engagement than the unstructured activities, as can be seen in Table 2.

Table 2. Mean Facial Engagement Proportion Across Various Activities ± SEM (SD)

Capture The Smile	Guess The Emotion	Unstructured Activity
0.466 ±0.032 (0.267)	0.366±0.023 (0.228)	0.280±0.015 (0.227)

5.4 Indicator Preferences

In order to obtain qualitative information about all four indicator options, families were asked to cycle through all four indicator options throughout the duration of the trial. During interviews, all 14 families stated when asked that they preferred the use of a facial indicator to not having one. One family remarked that the indicator wasn't displaying in certain lighting conditions outside, so they used it inside where the indicator confirmed operation was reliable. Another family commented that they would use the device often in the kitchen, because the indicator confirmed emotions could be detected there best.

This feedback supports our hypothesis that the indicator was an important aid in helping children distinguish between "neutral faces" and missed expression detections due to a system failure, such as those introduced by adverse lighting situations. Initially, we worried that this indicator would encourage the maladaptive behavior of children lining up the display with their conversation partner's face, but families reported no evidence of this, and no such behavior was observed in lab during study check-ins.

Of the four options for indicator shape and size presented to participants, parents and children chose the box indicator in slightly over 51% of all sessions, followed by the triangle indicator at 34%. This data further supports that children preferred having an indicator to just raw expression feedback and, given a visual design choice, chose the most prominent and noticeable one (the amount of face contact measured tended to decline as the area of the indicator declined). We observed a measurable difference in facial engagement across the four indicator choices, as seen in Figure 8. Facial engagement per indicator choice was calculated as the average over all sessions of the proportion of time the facial engagement indicator was activated in a session (weighted by session time). All visible indicator options provoked more facial engagement than having no indicator, with the largest gain measurable in the "box" indicator. However, although children volunteered during the interviews that they preferred the "triangle" indicator to the "line" indicator, the latter exhibited noticeably higher facial engagement. Nevertheless, the children responded that they preferred the "triangle" indicator to the line because "*it's easy to use*" (Participant 2, 10-year-old male) and "*it's not distracting*" (Participant 2, 10-year-old male as well as Participant 11, 4-year-old female) compared to the other indicator options. This discrepancy between some of the design preferences of the user and designs that result in higher facial engagement suggests that there may sometimes be a tradeoff between feedback that elicits a high level of facial contact versus feedback that is unobtrusive.
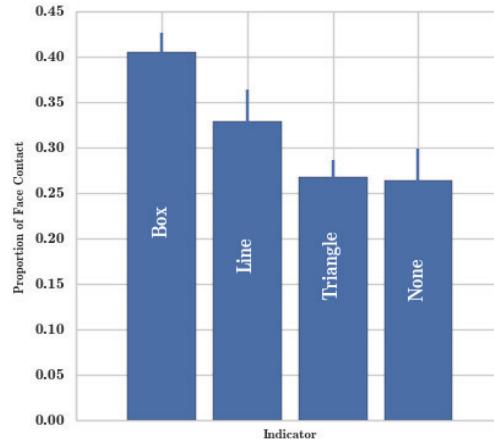
Fig. 8. Mean facial engagement for each indicator with SEM error-bars (t-test of "None" vs "Box" indicator gives p<0.05). Facial engagement per indicator choice was calculated as the average over all sessions of the proportion of time the facial engagement indicator was activated in a session (weighted by session time).

5.5 Parent Review

Based on usage logs, 7 out of 14 families used the parent review consistently more than once at home. During exit interviews, one family described using the review app to demonstrate their child's behavior to them, resulting in the child more carefully considering their behavior thereafter. The parent of another family came to realize, after seeing their child's field of view move sporadically, that their child's inability to focus was related to their frequent physical movements. One parent described the process of reviewing the session as a fun activity:

> "We like looking at the parent review app together. We usually wind up laughing at ourselves because, gosh, I didn't realize I looked like that!" – Participant 31 parent (participant 10 years old, male)

Regarding the families who did not use the parent review system, two families commented that the review's lack of audio playback (see Section 6.2) made it harder to review interactions because it was difficult to recall the context of the interaction from just an annotated video. Furthermore, we suspect that without building intentional activities centered around the parent review, families may not know how best to use it.

5.6 Model Accuracy

The base facial expression recognition system achieves 97.0% accuracy on the Extended Cohn-Kanade Dataset [22]. We performed cross-validation on our participant expression calibration data, achieving an accuracy of 74.7% (SD=10.5%) with participant-specific models and 60.4% (SD=10.5%) with the base model. Note that the much lower performance than on Cohn-Kanade owes to the fact that this data, gathered in an active learning fashion, is by definition "adversarial," with greater variation and noise than standard academic datasets, providing a lower bound estimate on real-time accuracy in the home.

While we maximized accuracy by customizing the emotion recognition model to each participant's family and close friends, there were still instances when the model failed. In some of these cases, parents were able to use the emotion misclassification as a learning experience:

"We would laugh if it picked up the wrong emotion and then we would talk about what is in fact the correct emotion...it gave us a platform to talk about emotions that we didn't have before." - Participant 2's Parent (participant 10 years old, male)

In other cases, when the emotion recognition failed, the parents would sometimes blame their emotional acting skills rather than the system itself:

"[Participant 17] has used the glasses several times with his therapists, and they seem to be working well. He's still getting inconsistent feedback on them when using them with me, though, which tells me that I really just don't have an expressive enough face for the software to be able to interpret my emotions correctly." – Participant 17's parent (participant 11 years old, male)

We are conducting further studies to evaluate the correlation between recognition accuracy and overall effectiveness of the device.

5.7 Potential to Increase Social Acuity

During the interviews, 11 out of 14 families responded in the exit interview that their child's ability to recognize emotions had noticeably improved during the course of the study. Participants noted:

"I've been applying what I saw with the Google Glass to situations without the Google Glass...I would see my friend's face and it would look similar to one of the faces I saw on my parents when they were upset so then I could ask my friend, 'What happened?'" - Participant 3 (15 years old, female)

"It's like she had to have a software upload and then she was able to implement it in her life and she just got it. She got the information she needed to start seeing how people were feeling and interacting" – Participant 18's parent (participant 11 years old, female)

This suggests that such wearable systems need not be prosthetics that are constantly worn by the children, but rather, can create improvement with limited and constrained daily use. Additionally, some parents also noticed that the activities involved in the study had caused them to communicate emotions more expressively, with more facial gesturing, than they typically had in the past. When asked in interviews, 13 out of 14 families responded in the exit interview that the study provided a context for family discussion and thereby had increased family quality time.

12 out of the 14 families that completed the study noticed a dramatic increase in eye contact, including cases where a difference was noticed independently by teachers and therapists who were unaware of the study. Some notable examples:

"We already noticed something very dramatically I like to share. [Participant 1] is actually looking at us when he talks through Google Glass during a conversation and it was noticed without glasses from his teacher in Language Art yesterday, its [sic] almost like a switch was turned. We found this very important to share and I hope it will help your research to take a closer look at this event when other kids will start wearing the glasses. Thank you!!! My son is looking into my face." - Participant 1's parent (participant 11 years old, male)

"I was at [Participant 17]'s school speaking with the office secretary there about something, and in reminiscing about how long [Participant 17] has been at the school she mentioned how [Participant 17] now looks her in the face when speaking to her (whereas previously [Participant 17] wouldn't look at her). I am seeing improved eye contact at home as well, though some days it's still hit or miss." - Participant 17's parent (participant 11 years old, male)

6 DISCUSSION

6.1 Lessons Learned

Based on our 12-week qualitative evaluation of the SuperpowerGlass system, we have learned several lessons regarding the design of wearable assistive technology that is used at home and controlled by parents. After the completion of the trials of all 14 participants, we reviewed both the yes/no answers to the interview questions, the supplemental parent and children quotes, system usage logs, and some session videos (described in Section 5.1) to identify key issues for our SuperpowerGlass system and for future designers of similar systems. Given the nature of this pilot design study with a small sample, we followed an iterative, inductive approach to finding patterns that pointed to potentially relevant outcomes and design issues. Members of our team reviewed the data together and identified areas where there appeared to be meaningful patterns to discuss.

Repeated Use of Wearable Social Aids Can Result in an Improvement in Social Behavior When Not Wearing the device. Seven of the 14 parents responded "yes" when asked if their child made increased social interaction when not wearing the device, and 11 of the 14 parents responded "yes" when asked if their child exhibited increased emotional recognition when not wearing the device. In addition to the improvements in social behavior noticed by most parents in the trial, we have seen various instances of people who were unaware of the SuperpowerGlass study, such as Participant 17's teacher, noticing improvements in child behavior (see Section 5.7). Future designers of wearable behavioral aids should consider whether the system needs to be continuously worn or whether an effect can be observed with minimal usage as was done in this study.

The Design of Wearable Systems For Children With ASD Should Be Heavily Dependent On the Functioning Level of the child. Since there is a significant discrepancy between children with high-functioning and low-functioning ASD with respect to activity preference, the development of the games should be targeted towards the child's age and severity of ASD symptoms. We also observed that certain children with ASD can handle multiple sensory stimulation. Both usage data and family interview feedback confirm that participants almost universally prefer enabling both a face indicator and video and audio feedback (12 of the 14 children preferred both audio and visual feedback, while 2 preferred visual only. None preferred audio only). This contradicts the initial concern that offering multiple kinds of feedback might cause sensory overload, and suggests that as we continue to develop the user experience, there may be room to experiment with increasing frequency and types of feedback.

In Order to Be engaging, Interactive Video Content from Wearable Therapy Sessions For Children With ASD Should Be Augmented With Context About the content. There was a lack of stated use of the post-session video review application despite frequent use of the Glass system (as described in Section 5.4). As discussed above, some parents found that the parent review feature helped facilitate engagement by providing a source of discussion and humor while others found little use for the app because it was difficult to recall the context of the interaction from a video only annotated with detected emotion. Parents stated that post-session tagging of content was rarely used for this reason. Therefore, interactive video content from wearable therapy sessions, when meant for review by children with ASD and their parents, should be augmented with sufficient context about the session so that users can easily navigate the video. This is particularly essential for long-term use of such wearable systems, which can produce hours of video content.

Limited Insight Into How the System Operates Can Make Users Doubtful of Their Own capabilities. When our emotion recognition model failed consistently for some parents, the parents concluded that this was due to their emotions not being expressive enough for detection. However, there are several reasons emotion recognition could fail. If the system were to expose more of how it operates, users would be aware that an emotion recognition failure may be due to other aspects such as poor lighting conditions or non-frontal head pose. In general, future designers should consider making the inner workings of automated

feedback mechanisms for therapeutic purposes transparent to the user, particularly for longitudinal use when failures could be corrected over time.

Children With ASD May Prefer System Design Decisions That Are Not Optimally Designed For the Desired Long-Term Outcome On the child. We noticed differences in facial engagement between the various indicator options. The amount of facial contact the children made with their parents increased as the area of the indicator increased, despite comments the children made of feeling more distracted with the larger facial indicators. Ideally, these secondary feedback cues can be helpful as long as they do not overwhelm the child. That is, the child should be able to focus on the primary content (such as the emotion cues) while simultaneously processing the peripheral information (face-tracking indicator light). Although children preferred the indicator that resulted in the most increase in facial engagement (the box) overall, for their second and third favorite choices, there was a stated discrepancy between design preferences of the user (the triangle indicator) and designs that resulted in higher facial engagement (the line indicator). This suggests that at times there can be a tradeoff between feedback that elicits a high level of facial contact versus feedback that is unobtrusive, and that the cost of two conflicting design goals (minimal obtrusiveness versus eliciting a response in the child) has to be weighed when determining the final design. While flashy designs have the potential to distract children with ASD from their primary task, in this case, the most obtrusive design was the one that resulted in the most effective clinical response (looking at a conversational partner's face). It is therefore important for designers to consider whether to prioritize outcome measures over participant preferences, particularly for participants with developmental challenges or delays.

6.2 Limitations

While the design trial provided several useful insights into the design of affective wearable systems for children with ASD, there were limitations of the study. The participants we recruited were all from a single geographical area that contains a large concentration of high-tech families. For example, 6 of the child participants were curious enough to actively attempt to access the native Android operating system, which is a nontrivial task because the application is set to be persistent. Furthermore, all but one of the study families had a yearly income greater than $100,000 per year. Because of the skewed population, it is not clear that the findings would generalize to a broader population.

The nature of our iterative trial resulted in some natural limitations in our results. During all participant check-ins and onboarding sessions, the most recent version of the Android and Glass software was uploaded to the device. Furthermore, the participants were recruited at different times, and therefore received differing versions of the system during onboarding. Specifically, the Guess the Emotion activity and line and triangle indicators were introduced shortly after enrolling the third participant. For analyses involving these choices, we only consider usage data from the first day that participants had all feature choices available to them. When excluding the first three affected participants from our plots, we show the same trends with some loss in statistical significance due to the decrease in sample size. Additionally, audio of activity sessions was originally not recorded due to privacy concerns. Audio recording for use in playback was enabled after confirming clearance with the Stanford Institutional Review Board, but was only available for a fraction of the study participants, so we did not include audio data in our analysis.

Further, the specific hardware available for the study also presented limitations to its implementation. The short battery life of Glass inherently limited some activity interaction design decisions. Also, while we believe the analysis method described in 5.3 to be a good proxy of facial engagement, availability of built-in or unobtrusive eye tracking capability would have allowed for a more accurate measure.

6.3 Ongoing Clinical Study

Given the inherently shifting nature of an iterative user design trial, our current study (which is well underway, although the data have not yet been examined) is a randomized controlled trial (RCT). We are keeping the implementation and protocol stable throughout the study.

The design study described in this paper has resulted in key design conclusions that have informed the blueprint for this RCT, such as:

- *Increasing family engagement.* We believe that additions in challenge and gamification, and in particular a reward system, will increase family engagement across the span of a study.
- *Standardization of therapy sessions.* The introduction of structure into the "Glass therapy sessions" conducted at home, partially administered by behavioral therapists on a regular basis. This will provide a basis for comparing facial engagement more consistently, and provide more consistency for frequency of device usage.
- *In-lab measures.* We are applying a consistently administered in-lab assessment to quantify eye contact to further reduce heterogeneity in facial engagement data.

6.4 Future Work

During the current clinical validation stage, ongoing development of the system is occurring in a parallel track to enable rapid prototyping and iteration while not interfering with the integrity of the RCT. For this ongoing development, we are exploring additional gamification, more reinforcing feedback, and activities that leverage the parent review. We are also exploring how audio analysis can further improve emotion recognition accuracy, and plan to experiment with other wearable devices, such as the Microsoft HoloLens.

Because the system tends to be used in interactions among a small group of people, we intend to explore ways that the emotion recognition models can be further improved by being iteratively trained by the parents on specific people. A mobile interface will allow parents to quickly train the emotion recognition model on themselves if certain emotions are frequently misclassified. Taking this further, we will use active learning techniques to query users with data gathered through their interaction with the device, prompting them to correct the classification if needed. Various active learning techniques intelligently choose key points for which user input is useful, minimizing the relabeling burden.

We would also like to explore and evaluate alternative games and activities that can be used with our system, including games that can subtly update the active learning model and those that can incorporate more complex emotions. The games could also include the addition of levels of complexity and features such as score tracking and reward systems.

7 CONCLUSIONS

The SuperpowerGlass system provides a powerful platform for providing real-time social cues to children with ASD. We have anecdotal evidence that the system has made a noticeable impact on the behavior of our participants even when not wearing the device. Design lessons from this system can be applied to other affective wearable systems in the future.

We have learned that 1) the repeated use of wearable social aids can result in an improvement in social behavior when not wearing the device, 2) children with ASD may prefer system design decisions that are not optimally designed for the desired long-term outcome on the child, 3) the design of wearable systems for children with ASD depends on the functioning level of the child, 4) limited insight into how the system operates can make users doubtful of their own capabilities, and 5) in order to be engaging, interactive video content from wearable therapy sessions for children with ASD should be augmented with context about the content.

REFERENCES

[1] Alex A. Ahmed and Matthew S. Goodwin. Automated Detection of Facial Expressions during Computer-Assisted Instruction in Individuals on the Autism Spectrum. 2017. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), 6050-6055.

[2] Clare Allen, Natalie Silove, Katrina Williams, and Paul Hutchins. 2007. Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders* 37,7: 1272–1278.

[3] Alberto Battocchi, Fabio Pianesi, Daniel Tomasini, Massimo Zancanaro, Gianluca Esposito, Paola Venuti, A. Ben Sasson, Eynat Gal, and Patrice L. Weiss. 2009. Collaborative Puzzle Game: a tabletop interactive game for fostering collaboration in children with Autism Spectrum Disorders (ASD). In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (ITS '09), 197-204.

[4] Kyle D. Bennett, Rangasamy Ramasamy, and Toby Honsberger. The effects of covert audio coaching on teaching clerical skills to adolescentswith autism spectrum disorder. 2013. Journal of Autism and Developmental Disorders 43,3: 585-593.

[5] Centers for Disease Control and Prevention and Centers for Disease Control and Prevention. 2014. Atlanta, GA: National Center on Birth Defects and Developmental Disabilities. Retrieved September 20, 2016 from http://www. cdc. gov/media/releases/2014/p0327-autism-spectrum-disorder.html.

[6] Chris J. Boyatzis and Reenu Varghese. 1994. Children's emotional associations with colors. The Journal of genetic psychology 155,1: 77-85.

[7] LouAnne E. Boyd, Alejandro Rangel, Helen Tomimbang, Andrea Conejo-Toledo, Kanika Patel, Monica Tentori, and Gillian R. Hayes. SayWAT: Augmenting Face-to-Face Conversations for Adults with Autism. 2016 In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 4872-4883.

[8] LouAnne E. Boyd, Xinlong Jiang, and Gillian R. Hayes. ProCom: Designing and Evaluating a Mobile and Wearable System to Support Proximity Awareness for People with Autism. 2017. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), 2865-2877.

[9] Navneet Dalal and Bill Triggs. 2005. Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), 886-893.

[10] Geraldine Dawson and Karen Burner. 2011. Behavioral interventions in children and adolescents with autism spectrum disorder: a review of recent findings. Current opinion in pediatrics 23,6: 616-620.

[11] Geraldine Dawson, Sally Rogers, Jeffrey Munson, Milani Smith, Jamie Winter, Jessica Greenson, Amy Donaldson, and Jennifer Varley. 2010. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. Pediatrics 125,1: e17-e23.

[12] Geraldine Dawson, Sara Jane Webb, and James McPartland. Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. Developmental neuropsychology 27,3: 403-424.

[13] Paul Ekman. 1992. An Argument for Basic Emotions. Cognition and Emotion 6,3-4: 169-200.

[14] Rana el Kaliouby, Rosalind Picard, and Simon Baron-Cohen. 2006. Affective computing and autism. Annals of the New York Academy of Sciences 1093,1: 228-248.

[15] Lizbeth Escobedo, David H. Nguyen, LouAnne Boyd, Sen Hirano, Alejandro Rangel, Daniel Garcia-Rosas, Monica Tentori, and Gillian Hayes. 2012. MOSOCO: a mobile assistive tool to support children with autism practicing social skills in real-life situations. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), 2589-2598.

[16] Lizbeth Escobedo, Monica Tentori, Eduardo Quintana, Jesus Favela, and Daniel Garcia-Rosas. 2014. Using augmented reality to help children with autism stay focused. IEEE Pervasive Computing 13,1: 38-46.

[17] David Feil-Seifer and Maja J. Matarić. 2009. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In Experimental robotics: 201-210.

[18] Jenny Rose Finkel and Christopher D. Manning. 2009. Hierarchical bayesian domain adaptation. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 602-610). Association for Computational Linguistics.

[19] Eynat Gal, Nirit Bauminger, Dina Goren-Bar, Fabio Pianesi, Oliviero Stock, Massimo Zancanaro, and Patrice L. Tamar Weiss. 2009. Enhancing social communication of children with high-functioning autism through a co-located interface. Ai & Society 24,1: 75-84.

[20] Leonardo Giusti, Massimo Zancanaro, Eynat Gal, and Patrice L. Tamar Weiss. 2011. Dimensions of collaboration on a tabletop interface for children with autism spectrum disorder. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11), 3295-3304.

[21] Ofer Golan, Emma Ashwin, Yael Granader, Suzy McClintock, Kate Day, Victoria Leggett, and Simon Baron-Cohen. 2010. Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. Journal of autism and developmental disorders 40,3: 269-279.

[22] Nick Haber, Catalin Voss, Azar Fazel, Terry Winograd, and Dennis Wall. A practical approach to real-time neutral subtraction for facial expression recognition. IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.

[23] Gillian R. Hayes, Julie A. Kientz, Khai N. Truong, David R. White, Gregory D. Abowd, and Trevor Pering. 2004. Designing capture applications to support the education of children with autism. In International Conference on Ubiquitous Computing (Ubicomp '14), 161-178.

[24] Gillian R. Hayes, Sen Hirano, Gabriela Marcu, Mohamad Monibi, David H. Nguyen, and Michael Yeganyan. 2010. Interactive visual supports for children with autism. Personal and ubiquitous computing 14,7: 663-680.

[25] Javier Hernandez and Rosalind W. Picard. 2014. SenseGlass: using google glass to sense daily emotions. In Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology (UIST '14), 77-78.

[26] Sen H. Hirano, Michael T. Yeganyan, Gabriela Marcu, David H. Nguyen, Lou Anne Boyd, and Gillian R. Hayes. 2010. vSked: evaluation of a system to support classroom activities for children with autism. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10), 1633-1642.

[27] Patricia Howlin, Susan Goode, Jane Hutton, and Michael Rutter. 2004. Adult outcome for children with autism."Journal of Child Psychology and Psychiatry 45,2: 212-229.

[28] Warren Jones and Ami Klin. 2013. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. Nature 504,7480: 427-431.

[29] Michelle R. Kandalaft, Nyaz Didehbani, Daniel C. Krawczyk, Tandra T. Allen, and Sandra B. Chapman. 2013. Virtual reality social cognition training for young adults with high-functioning autism. Journal of autism and developmental disorders 43,1: 34-44.

[30] Naz Kaya and Helen H. Epps. 2004. Color-emotion associations: Past experience and personal preference. In Proceedings of Color and Paints, Interim Meeting of the International Color Association 5,31.

[31] Marion Koelle, Matthias Kranz, and Andreas Möller. 2015. Don't look at me that way!: Understanding User Attitudes Towards Data Glasses Usage. In Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15), 362-372.

[32] Rebecca J. Landa, Katherine C. Holman, and Elizabeth Garrett-Mayer. 2007. Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. Archives of general psychiatry 64,7: 853-864.

[33] Rebecca J. Landa, Katherine C. Holman, Allison H. O'Neill, and Elizabeth A. Stuart. 2011. Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: a randomized controlled trial. Journal of Child Psychology and Psychiatry 52,1: 13-21.

[34] Ole I. Lovaas. Behavioral treatment and normal educational and intellectual functioning in young autistic children. 1987. Journal of consulting and clinical psychology 55,1: 3.

[35] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. 2010. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops (CVPR '10), 94-101.

[36] Michael Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. 1998. Coding facial expressions with gabor wavelets. In Proceedings of the third IEEE International Conference on Automatic Face and Gesture Recognition (FG '98), 200-205.

[37] Gabriela Marcu, Anind K. Dey, and Sara Kiesler. 2012. Parent-driven use of wearable cameras for autism support: a field study with families. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (Ubicomp '12), 401-410.

[38] James C. McPartland, Brian Reichow, and Fred R. Volkmar. 2012. Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. Journal of the American Academy of Child & Adolescent Psychiatry 51,4: 368-383.

[39] David Moore, Yufang Cheng, Paul McGrath, and Norman J. Powell. 2005. Collaborative virtual environment technology for people with autism. Focus on Autism and Other Developmental Disabilities 20,4: 231-243.

[40] Brenda Smith Myles, Winnie Dunn, Louann Rinner, Taku Hagiwara, Matthew Reese, Abby Huggins, and Stephanie Becker. 2004. Sensory issues in children with Asperger syndrome and autism. Education and training in developmental disabilities: 283-290.

[41] Letizia Palumbo, Hollie G. Burnett, and Tjeerd Jellema. 2015. Atypical emotional anticipation in high-functioning autism. Molecular autism 6,1: 1.

[42] Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat. 2005. Web-based database for facial expression analysis. In 2005 IEEE international conference on multimedia and Expo.

[43] Ji Hyun Park, Bretagne Abirached, and Yan Zhang. 2012. A framework for designing assistive technologies for teaching children with ASDs emotions. In CHI' 12 Extended Abstracts on Human Factors in Computing Systems (CHI '12), 2423-2428).

[44] Rosalind W. Picard. 2010. Affective computing: from laughter to IEEE. IEEE Transactions on Affective Computing 1,1: 11-17.

[45] Kaska Porayska-Pomsta, Christopher Frauenberger, Helen Pain, Gnanathusharan Rajendran, Tim Smith, Rachel Menzies, Mary Ellen Foster, Alyssa M Alcorn, S Wass, S Bernadini, Katerina Avramides, Wendy Keay-Bright, Jingying Chen, A Waller, Karen Guldberg, Judith Good, and Oliver Lemon. Developing technology for autism: an interdisciplinary approach. Personal and Ubiquitous Computing 16,2 (2012): 117-127.

[46] Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K. Kane. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16), 4884-4895.

[47] Gale H. Roid. 2003. Stanford–Binet intelligence scales (5th ed.): Technical manual. Itasca, IL: Riverside Publishing.

[48] Noah J. Sasson, Amy E. Pinkham, Lauren P. Weittenhiller, Daniel J. Faso, and Claire Simpson. 2015. Context effects on facial affect recognition in schizophrenia and autism: behavioral and eye-tracking evidence. Schizophrenia bulletin: sbv176.

[49] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. 2008. Bosphorus database for 3D face analysis. In European Workshop on Biometrics and Identity Management: 47-56.

[50] Brian Scassellati, Henny Admoni, and Maja Mataric. 2012. Robots for use in autism research. Annual review of biomedical engineering 14: 275-294.

[51] Terence Sim, Simon Baker, and Maan Bsat. 2002. The CMU pose, illumination, and expression (PIE) database. In Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FG '02), 46-51.

[52] Katharina Spiel, Christopher Frauenberger, Eva Hornecker, and Geraldine Fitzpatrick. When Empathy Is Not Enough: Assessing the Experiences of Autistic Children with Technologies. 2017. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), 2853-2864.

[53] Katharina Spiel, Laura Malinverni, Judith Good, and Christopher Frauenberger. Participatory Evaluation with Autistic Children. 2017. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17), 5755-5766.

[54] Xiaoyang Tan and Bill Triggs. 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE transactions on image processing 19,6: 1635-1650.

[55] Alea C. Teeters. 2007. Use of a wearable camera system in conversation: Toward a companion tool for social-emotional learning in autism. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

[56] Patricia Valdez and Albert Mehrabian. 1994. Effects of color on emotions. Journal of experimental psychology: General 123,4: 394.

[57] Catalin Voss, Peter Washington, Nick Haber, Aaron Kline, Jena Daniels, Azar Fazel, Titas De, Beth McCarthy, Carl Feinstein, Terry Winograd, and Dennis Wall. 2016. Superpower glass: delivering unobtrusive real-time social cues in wearable systems. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (Ubicomp '16), 1218-1226.

[58] Peter Washington, Catalin Voss, Nick Haber, Serena Tanaka, Jena Daniels, Carl Feinstein, Terry Winograd, and Dennis Wall. 2016. A Wearable Social Interaction Aid for Children with Autism. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI '16), 2348-2354.

[59] Kristin Williams, Karyn Moffatt, Denise McCall, and Leah Findlater. 2015. Designing conversation cues on a head-worn display to support persons with aphasia. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), 231-240.

[60] Jean Xavier, Violaine Vignaud, Rosa Ruggiero, Nicolas Bodeau, David Cohen, and Laurence Chaby. 2015. A Multidimensional Approach to the Study of Emotion Recognition in Autism Spectrum Disorders. Frontiers in psychology 6.

[61] Michael T. Yeganyan, Meg Cramer, Lou Anne Boyd, and Gillian R. Hayes. 2010. vSked: an interactive visual schedule system for use in classrooms for children with autism. In Proceedings of the 9th International Conference on Interaction Design and Children (IDC '10), 319-322.

[62] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. 2008. A high-resolution 3D dynamic facial expression database. In 8th Proceedings on the IEEE International Conference On Automatic Face & Gesture Recognition (FG'08), 1-6.