Feasibility Testing of a Wearable Behavioral Aid for Social Learning in Children with Autism

Jena Daniels[1]   Nick Haber[1]   Catalin Voss[1,2]   Jessey Schwartz[1]   Serena Tamura[3]   Azar Fazel[1]
Aaron Kline[1]   Peter Washington[1]   Jennifer Phillips[3]   Terry Winograd[2]   Carl Feinstein[3]   Dennis P. Wall[1,4]

[1] Division of Systems Medicine, Department of Pediatrics, Stanford University, California, United States
[2] Department of Computer Science, Stanford University, California, United States
[3] Department of Psychiatry and Behavioral Sciences, Stanford University, California, United States
[4] Department of Biomedical Data Science, Stanford University, California, United States

Address for correspondence Dennis P. Wall, PhD, Division of Systems Medicine, Department of Pediatrics, Department of Biomedical Data Science, Stanford University, 1265 Welch Road, Suite X143, Stanford, CA 94305, United States (e-mail: dpwall@stanford.edu).

Abstract

Background   Recent advances in computer vision and wearable technology have created an opportunity to introduce mobile therapy systems for autism spectrum disorders (ASD) that can respond to the increasing demand for therapeutic interventions; however, feasibility questions must be answered first.

Objective   We studied the feasibility of a prototype therapeutic tool for children with ASD using Google Glass, examining whether children with ASD would wear such a device, if providing the emotion classification will improve emotion recognition, and how emotion recognition differs between ASD participants and neurotypical controls (NC).

Methods   We ran a controlled laboratory experiment with 43 children: 23 with ASD and 20 NC. Children identified static facial images on a computer screen with one of 7 emotions in 3 successive batches: the first with no information about emotion provided to the child, the second with the correct classification from the Glass labeling the emotion, and the third again without emotion information. We then trained a logistic regression classifier on the emotion confusion matrices generated by the two information-free batches to predict ASD versus NC.

Results   All 43 children were comfortable wearing the Glass. ASD and NC participants who completed the computer task with Glass providing audible emotion labeling ($n = 33$) showed increased accuracies in emotion labeling, and the logistic regression classifier achieved an accuracy of 72.7%. Further analysis suggests that the ability to recognize surprise, fear, and neutrality may distinguish ASD cases from NC.

Conclusion   This feasibility study supports the utility of a wearable device for social affective learning in ASD children and demonstrates subtle differences in how ASD and NC children perform on an emotion recognition task.

Keywords
► autism spectrum disorder
► pilot projects
► wearable device and body area network
► artificial intelligence
► emotion recognition
► pediatrics
► people with disabilities or special needs

Background and Significance

More than 3.5 million people in the United States are on the autism spectrum.[1] Autism is the fastest-growing developmental disability in the United States; within 5 years, prevalence rates of autism spectrum disorder (ASD) have climbed from 1 in 125 to 1 in 68 U.S. children, and the overall estimated prevalence is 14.6 children aged 8 years old per 1,000. [2–5] Most of these children struggle to make eye contact, recognize facial expressions, and engage in social interactions.[6–12] Impairments in facial affect recognition may contribute to social disability in autism, and improving

this skill is important for greater social development.[11,13] Additionally, emotional empathy is positively correlated with expression recognition ability in typically developing children.[14–16] To teach such social skills, behavioral intervention therapies for autism today, such as applied behavioral analysis (ABA) and Naturalistic Developmental Behavioral Interventions (NDBIs), involve social interactions with certified clinicians, which can pose challenges with generalizability, social anxiety, and understanding pragmatics.[17,18] Furthermore, with the rate of ASD diagnosis increasing, clinical centers have become outnumbered by, and out of reach from, the many children and families in need of attention. Waitlists for access are up to 18 months long, oftentimes delaying the onset of therapy until after developmentally sensitive periods have passed.[19,20] Consequently, many children with autism fail to build core social skills and subsequently regress down a path of isolation that worsens their long-term prognosis.[21–23]

Parsons et al, Moore and Calvert, and Golan et al support the use of computers and software as an effective platform for teaching skills to children with ASD due to its predictable nature, removing problems with pragmatics and anxiety associated with social interactions.[24–30] Furthermore, technology enables engagement without direct human interaction. Videos, photographs, and voice recordings are often used to explain concepts and tasks. However, by replacing human interactions with technology, a more dynamic and personalized experience is often lost. To incorporate real-time social interactions with the use of wearable technologies for children with ASD, Madsen et al created a mobile PC program that creates emotion "bubbles" indicating the emotions of others interacting with a person with ASD in real time.[31] Similarly, Liu et al created a system using Google Glass and face detection software to create structured games for children with autism during real-time interactions.[32] These projects highlight the importance of real-time social interactions for technology-based therapeutic interventions.

We hypothesize that the use of mobile technology to aid children with autism during natural social interactions can improve both the quality of and access to therapy. We have built a prototype[33] of such a mobile system on Google Glass, which confers a form of augmented reality that can deliver therapeutic information, such as the emotion in a person's face, without immersing the wearer in a virtual world removed from their natural environment.

While our ultimate goal is to understand this tool's potential for delivering home therapy, our first step was to test the feasibility of the system in a controlled laboratory setting on children with and without ASD. Individuals with ASD benefit from behavioral interventions involving social interactions for affect training,[7,8,18] but how and why this population perceives expressions differently from neurotypical control (NC) is not entirely understood.[6,18,22,34,35] Ozonoff et al and Castelli et al have reported little disparity in facial affect recognition between children with ASD versus children who are NC.[36,37] These studies suggest that for individuals with ASD, the ability to recognize emotions is the same as (or at least similar to) NC, despite the tendency

that they are less interested in social interactions.[37] Even though children with ASD may have the same ability to recognize facial expressions, studies have shown subtle but significant differences while processing these faces, such as a lack of eye fixation with more fixation on mouths,[38,39] and a longer processing time to analyze faces.[34,39] Furthermore, some studies have found ASD deficits in recognizing certain emotions more often than others, such as happiness and neutrality,[40] surprise,[36,38] and fear,[36,39,40] as well as deficits in judgment of more complex social stimuli, such as trustworthiness, shame, and approachability.[34,36,38,41]

We outline our hypotheses and objectives below.

Objective

To our knowledge, no system using wearable technology has yet taught facial affect recognition to children with autism, even though training in facial affect recognition has been shown to be highly beneficial for children with autism.[38,42–44] We strive to create such a system. However, as a first step in the direction of testing the above system and to understand the feasibility of using Google Glass as a platform for mobile-therapeutic intervention, we sought to confirm the following hypotheses in this study:

> **Hypothesis 1:** Such a device and the visual and/or audio information about emotion given to children with ASD will be comfortable (not overstimulating for sensory sensitivities).
> **Hypothesis 2:** The information provided to the child by the Glass system in the form of visual and audio information will be an effective way to increase facial affect recognition skills.
> **Hypothesis 3:** Children with ASD and the NC group differ in their abilities to recognize facial expressions.

To test these hypotheses, we created a software capable of running on Google Glass and designed to give preprogrammed emotion-feedback cues to the wearer. We then explored the hypotheses above in a controlled laboratory setting. Each participant was asked to recognize static facial expressions displayed on a computer screen in three separate batches, with the middle batch giving the correct emotion-feedback cue (i.e., corresponding to the true value of the facial expression displayed on the static image). As most children with ASD have altered sensitivities to sensory input,[45–49] this feasibility study allowed us to ensure that Google Glass was not overstimulating, as well as ensure the method of delivering such emotion-feedback was developmentally appropriate. We then analyzed the difference in response patterns between ASD and NC participants.

Methods

Participant Recruitment
Between February 25, 2015 and January 26, 2016, we enrolled 43 participants under an approved Institutional Review Board (IRB) protocol at Stanford University. Of those $n = 23$ had a confirmed clinical diagnosis of ASD and $n = 20$ were NC. The ASD cohort was 82.61% male ($n = 19$ male, $n = 4$ female) with

an average age of 11.65 years (minimum age = 6 years, maximum age = 17 years, SD = 3.20), while 70% of the NC were male ($n = 14$ male, $n = 6$ female) with an average age of 11.55 years (minimum age = 7 years, maximum age = 17 years, SD = 3.09). Participants were identified from referral to the Autism and Developmental Disabilities Clinic and the Developmental Behavioral Unit of Lucile Packard Children's Hospital, as well as from academic presentations and clinical services by the Stanford faculty and staff on the study team. Additionally, participants were identified from the Autism and Developmental Disabilities Research Registry.

Eligible candidates were identified through a qualitative survey, review of medical history, and then by the Social Communication Questionnaire (SCQ[50]) via the phone. ASD diagnosis was confirmed by the SCQ and later by a copy of the participant's clinical medical diagnosis. An expert review by a licensed psychologist was completed for each subject to confirm ASD classification. Participants with ASD were not included if: (1) they had evidence of a genetic, metabolic, or infectious etiology for their autism based on medical history; (2) had a history of seizures or other neurological problems; (3) required assistance with their vision; and/or (4) they had a diagnosis of any severe mental disorder such as schizophrenia or bipolar disorder.

NC participants were excluded if: (1) > 14 score on the SCQ; (2) a diagnosis of a neurological disorder currently or in the past on the basis of self-report and medical history; (3) a diagnosis of psychiatric disorders currently and in the past on the basis of a clinical psychiatric evaluation and information obtained from

behavioral scales; (4) had a sibling diagnosed with ASD or schizophrenia or had evidence of ASD or schizophrenia; (5) they had a history of seizures or other neurological problems; and/or (6) they required assistance with their vision.

Laboratory Testing

Participants and parents provided written informed consent under an approved Stanford University IRB protocol, which followed the guidelines of the Declaration of Helsinki,[51] prior to their inclusion in the study. During the study appointment, a trained research assistant assessed each participant with the Stanford Binet Intelligence Scales, Fifth Edition, Abbreviated Battery Intelligence Quotient (ABIQ[52]). In addition, we collected the parent-completed Social Responsiveness Scale-2 (SRS-2[53]).

Computer Task

During the computer task, participants wore Google Glass and were asked to recognize the emotion from a static image displayed on a computer monitor (►Fig. 1). Participants were asked to classify 125 static facial images of children on a computer screen that were broken up into three rounds of successive "batches" of images. All images were drawn from a standard data set[11,54] of clear facial expressions, as one of seven universal facial expressions: Happy, Sad, Angry, Scared, Disgust, Surprised, and Calm.[55] Selected images were expression-, race-, and gender-balanced. To both the left and right of each image, nonsocial distraction stimuli were displayed (selected from a peer-reviewed data set[56]) to ensure relevant stimuli were not shown in isolation. The static facial image and
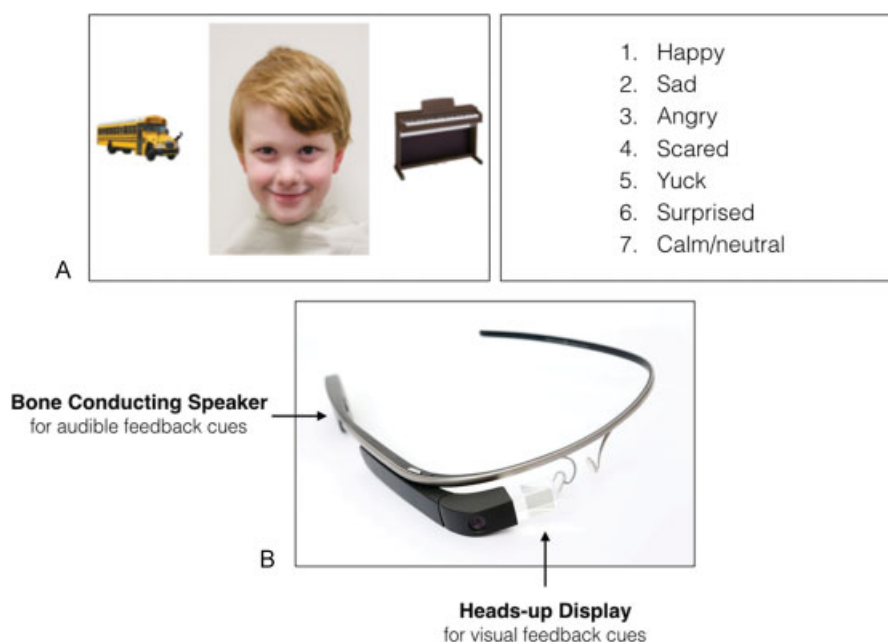

Fig. 1 Example of a static emotion and response list provided to participants during the computer task and depiction of the Glass. (**A**) Sample image of computer monitor display during the computer task. In batches 1 and 3, the computer monitor displayed an image of a child expressing one of seven emotions (Happy, Sad, Angry, Scared, Yuck, Surprised, Calm/neutral), similar to the one on the left, for 6 seconds, and then automatically displayed the screen on the right with the various options for as long as it took for the participant to decide the emotion expressed from the image displayed on the previous screen. In batch 2, the image on the left was still displayed for 6 seconds; however, after 3 of the 6 seconds while the image was displayed, the Glass would provide the feedback cue to the participant before transitioning to the screen on the right. (**B**) Sample depiction of the Glass that participants wore during study procedures, which shows the heads-up display for visual cues and the bone-conducting speaker for audio cues.

the nonsocial stimuli were displayed for 6 seconds followed by a list of the seven possible emotions (Happy, Sad, Angry, Scared, Yuck, Surprised, and Calm). The list was displayed until the participant chose a response.

In the first and third batches of images during the computer task, participants received no information from the Glass during the 6 seconds each image was displayed on the computer monitor. In the second batch, although each image was still displayed for 6 seconds total, after 3 seconds of displaying the image, the Glass provided the correct labeled emotion for the displayed image. The sessions were all video recorded and a researcher accompanied the participant through the task, recording the response option. Responses were double confirmed via the video session recording.

Glass Feedback

Our first four participants (all NC) received only visual information on the Google Glass heads-up display in the form of the word corresponding to the correct expression. Upon early analysis and qualitative feedback from these participants who received visual information only, we determined that some of the participants were unable to read the visual cues and hence their data were not included in the analyses for Hypotheses 2 and 3 (refer to the Results sections Hypotheses 2 and 3). We also learned that most children had difficulty reading and verbalizing the word "disgusted," and therefore

for the purposes of the study, we referred to "disgusted" as "yuck" while working with participants. Additionally, the participants found it challenging to read the word "surprised" from Google Glass heads-up display window. As "surprised" and "disgusted" were the longest words and the font size on Google Glass heads-up display was small, we learned that children had a hard time reading these words from the heads-up display but did not have any difficulties when reading these words in the visual prompt provided to them during the computer task. In audiovisual speech integration studies for adolescents with ASD, DePape et al and Smith and Bennetto found that there was no benefit to providing both visual and audio information, in comparison to providing only audio cues during speech perception tasks.[57,58] For these reasons, audio feedback delivered using the Glass's bone-conducting output speakers (►Fig. 1B) in the form of the word corresponding to the correct expression in the image, was determined to be the most appropriate method for providing information to children for this in-laboratory study.

Participant Demographics

Forty-three participants qualified for Hypothesis 1, confirming that they were able to comfortably wear Google Glass. We excluded 10 of these 43 participants from Hypotheses 2 and 3 due to technical or study procedure completion issues (►Fig. 2). With the first four excluded participants (all NC
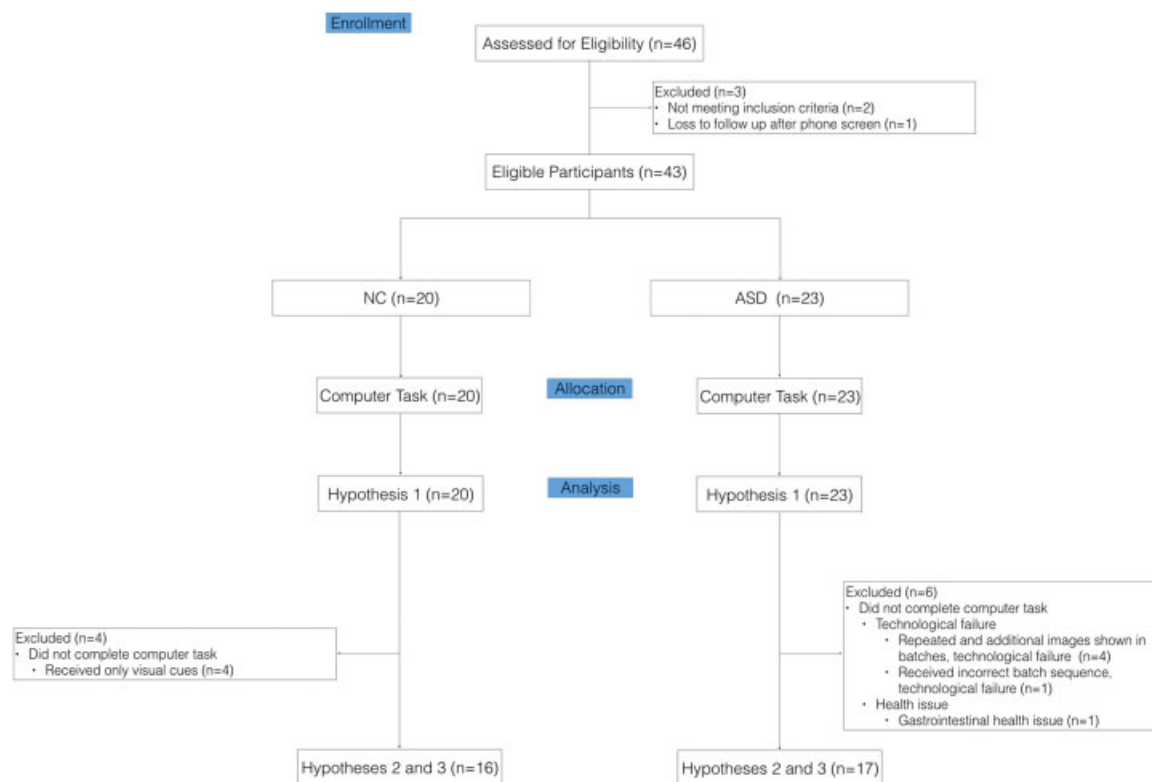

Fig. 2 Consort flow diagram. We assessed 46 interested participants for eligibility. Forty-three participants met inclusion requirements for the study, 20 neurotypical controls (NCs) and 23 children with autism spectrum disorder (ASD). We allocated all 43 eligible participants the computer task and all 43 qualified for testing Hypothesis 1. We excluded $n = 4$ NC and $n = 6$ ASD for testing of Hypotheses 2 and 3. The six excluded ASD participants experienced either a technological failure in the software or could not proceed with the study due to health. In addition, we excluded the first four participants (all of whom were neurotypically developing) due to a change in the study procedures following their participation. Namely, these first four subjects struggled with the visual feedback (written words on the Glass units' heads-up display) compelling us to use only audio cues for the remaining participants and for the remaining duration of the study.

males ages 7–11 years with ABIQs ranging from 85 to 106 [SD = 8.83]), we attempted to use visual cues in the form of written words on the Glass units' heads-up display (►Fig. 1B). However, these children struggled to read the words on the display. Therefore, we adapted our design to use only audio cues for the remaining subjects. Because this made the data from these original four incomparable to the data from the majority of the subjects, we elected to exclude the four from use in testing Hypotheses 2 and 3. Additionally, we excluded six ASD participants from tests of Hypotheses 2 and 3, five males and one female, all between 6 and 11 years of age who scored between 85 and 115 (SD = 9.98) on the ABIQ.[52] In five of the six cases, there was a technological error, displaying incorrect images during a batch in the computer task. In the last case, the child experienced a gastrointestinal health issue and was unable to complete the computer task.

Our results for Hypotheses 2 and 3 are based on the remaining 33 participants ($n = 16$ ASD, $n = 17$ NC). The average age was 12.13 years for the ASD cohort (minimum age = 6, maximum age = 17, SD = 3.31 years), and 81.25% were male ($n = 13$ males, $n = 3$ females). The average age was 11.53 for the NC (minimum age = 8, maximum age = 17, SD = 2.48 years), and 52.94% were male ($n = 9$ males, $n = 8$ females). The average ABIQ was 102.75 (minimum = 55, maximum = 133, SD = 19.54) for the ASD cohort and 108.94 (minimum = 91, maximum = 129, SD = 9.58) for NC. NC had an average SCQ score of 1.82 (minimum score = 0, maximum score = 4, SD = 1.07), while the average SCQ for the ASD cohort was 18.86 (minimum score = 7, maximum score = 31, SD = 6.43). The mean SRS t-score for the ASD group was 67.54 (minimum = 52, maximum = > 90, SD = 11.13). For NC, the mean SRS t-score was 44.12 (minimum = 30, maximum = 64, SD = 8.11). Of the 17 NC, 16 scored a "normal" (< 59) SRS-2 severity rank, and one scored a "mild" (60–65) severity rank. Ten of the 16 ASD participants scored a "severe" (> 76) SRS-2 severity rank, one scored a "moderate" (65–75) severity rank, one scored a "mild" (60–65) severity rank, and one scored a "normal" (< 59) severity rank. All ASD participants provided a clinical diagnostic report, confirming his/her professional ASD diag-nosis. See ►Table 1 for the mean and standard deviation for SRS-2 t-scores for our ASD cohort versus NC.

Data Analysis

In what follows, we provide an analysis of the performance of the participants on the expression recognition task, starting first with an analysis of accuracies before moving to an analysis of the participants' confusion between expressions.

Hypothesis 2

We analyzed our data for Hypothesis 2 using a two-way repeated measures analysis of variance (ANOVA) analysis with percent accuracy during the facial affect recognition task as a repeated measure and batch as a grouping factor. We examined both the ASD and NC groups in this way.

Hypothesis 3

Although both the ASD and NC groups performed similarly well in the emotion recognition test, they differed in which emotions they mislabeled. We computed the confusion matrices for each batch of each participant, defined to be the square matrix with rows and columns corresponding to the possible responses, with the entry in row r and column c to be the number of images for which the true value corresponds to r but the participant gave answer corresponding to c. We order the expressions as follows: Happy, Angry, Surprise, Neutral, Fear, Sad, and Disgust. The rows are the true emotions and the columns are the ones indicated by the child, displaying the frequency with which the child mixed emotions for all pairs of emotions possible. Hence, if a participant's batch 1 confusion matrix has a 3 in the third row and fifth column, the participant mistook three surprise faces for fear faces in the first batch.

We then evaluated how well a machine-learning algorithm can predict the participant's diagnosis (ASD or NC) based on their confusion matrices. It should be cautioned that our analysis is based on a small amount of data. It is also difficult, even with proper cross-validation executed, to choose an appropriate model and avoid overfitting. We chose logistic regression with L_1 regularization, because the regularization term not only fights against overfitting, but also

Table 1 Parent-reported Social Responsiveness Scale (SRS) t-scores for ASD and NC cohorts

SRS subdomain	Mean t-scores (SD)		Significance
	NC	ASD	
Total SRS	44.1 (8.11)	78.8 (11.13)	> 0.001
Social awareness	46.2 (9.93)	67.5 (12.64)	> 0.001
Social cognition	44.4 (7.00)	73.6 (13.46)	> 0.001
Social communication	47.0 (10.19)	74.2 (12.16)	> 0.001
Social motivation	44.4 (8.07)	69.2 (14.56)	> 0.001
Social mannerisms	44.1 (3.66)	76.3 (10.20)	> 0.001

Abbreviations: ASD, autism spectrum disorder; NC, neurotypical control; SD, standard deviation.
Note: A higher SRS score indicates more severe ASD symptoms. An independent samples t-test indicates significant differences between NC and ASD on total SRS t-scores and subdomain scores.

enforces sparsity, favoring models that pay attention to fewer elements of the confusion matrices.[59] Lacking the data to make further refinements in model selection, this is a reasonable first choice of a simple, linear model given the type of data.

We performed nested cross-validation, with the outer nest made in a leave-one-out fashion, and with the inner 10-fold stratified validation for tuning the regularization parameter. Confusion matrices were normalized so that their rows summed to 1 as a preprocessing step to this classification task.

To understand the patterns these classifiers exploit, we look at the coefficient the logistic regression assigns to each entry in the confusion matrices, training one model on all the data using the most frequently chosen regularization parameter. Logistic regression, when input a feature x, learns weights w and bias b and bases its decision on the sign of

$$s = (w\hat{\ }t)\, x + b$$

with a positive decision (in this case ASD) if s is positive and a negative decision (in this case NC) if s is negative (further probabilistic interpretation is based on a sigmoid function applied to s). This can give us a sense of what the classifier deems "more ASD" or "more NC."

Results

Hypothesis 1
All 43 participants amenably wore Google Glass for at least 15 minutes and qualified for use in testing of the hypothesis that Google Glass is feasible for use in children with autism needing behavioral therapy. The majority of participants commented that they were intrigued by the hardware. Participants either received only visual cues ($n = 4$) on Google Glass' heads-up display as words (Happy, Sad, Angry, Scared, Surprised, Yuck, or Calm) or only audio cues ($n = 39$) via the bone-conducting speaker on the Glass. Based on the comfort level, length of time the Glass was worn, and qualitative feedback provided by participants, we determined that Google Glass is a feasible and wearable hardware solution to provide unobtrusive social cues to children with ASD. Ten could not be included in testing of Hypotheses 2 and 3, either due to receiving visual cues via the Glass ($n_{NC} = 4$), a gastrointestinal health issue ($n_{ASD} = 1$), or technological failures ($n_{ASD} = 5$: one received the incorrect batch sequence and four received additional or repeated images during a batch) (►Fig. 2).

Hypothesis 2
Both children with ASD and NC improved from batch 1 to batch 3. In batch 1, both groups incorrectly labeled ~15.8% of the images. Children with ASD did not differ significantly ($p = 0.927$) from NC in their accuracy in labeling the static facial images with emotion. From the classification information delivery enabled by the Glass in batch 2, this frequency of incorrect labeling dropped to 6.1% in the autism cohort and 5.2% in the NC cohort ($p = 0.732$). After emotion training and returning to the information-free mode in batch 3, both groups showed a sustained increase in correct labeling of emotions, with a significant increase of NC by 5.2% ($p = 0.011$) and a slight increase in accuracy for the ASD cohort by 2.6% ($p = 0.058$). Although we did not provide a control group during this study to compare children who received no information from the Glass and therefore we cannot assume a learning effect, receiving emotion classification information from the Glass was associated with a sustained increase in accuracy for both ASD and NC during a facial affect recognition task, supporting Hypothesis 2. See ►Table 2 for overall accuracies of participants by batch and ►Fig. 3 for a graph of the results. This difference was not statistically significant.

Hypothesis 3
For the two batches without information regarding emotion from the Glass units, we achieved validation accuracy of 69.7% ($p = 0.047$ by the permutation test[60]) and 72.7% ($p = 0.036$), respectively. For the batch with emotion information provided to the child, we achieved validation accuracy of 51.5%, the level of accuracy one can achieve by always classifying the participant as an NC—the algorithm was unable to find a difference between ASD performance and NC performance in this batch.

See ►Tables 3 and 4 for the results on batches 1 and 3, respectively. Weights have been arranged in the shape of the confusion matrix, so, for instance, in the case of ►Table 4, a weight of –2.16 in the Fear row and Surprise column gets multiplied by the confusion matrix entry corresponding to a true value of Fear and a response of Surprise. Intuitively, positive entries indicate that the algorithm interprets larger confusion matrix values in that location as "more likely ASD," while negative entries correspond to "more likely neurotypical."(We caution against assigning relative size of coefficients too much meaning; for training, features [participants' confusion matrices] are scaled so that each row sums to 1. Hence, some entries [off-diagonal] are expected to be smaller than other entries [on-diagonal]. Ordinarily, one would rescale features to have mean 0 and variance 1, or minimum 0 and maximum 1, which would make the magnitudes easier to compare, but requires an understanding of the scale of each feature in addition to the classifier's coefficients; the amount of data relative to the number of features made this infeasible.)

Table 2 Overall accuracies of participants by batch

Overall accuracies of participants by batch			
	Batch 1	Batch 2	Batch 3
ASD	85.3%	93.9%	87.9%*
NC	85.1%	94.8%	90.3%**

Abbreviations: ASD, autism spectrum disorder; NC, neurotypical control.
Note: Two asterisks indicate the significance in change in accuracy between batches 1 and 3 (**$p < 0.05$). One asterisk (*) indicates a marginally significant change in accuracy. The first batch received no information from the Glass, the second batch received classification information as audio, visuals, or both from the Glass, and the third batch again received no information from the Glass.
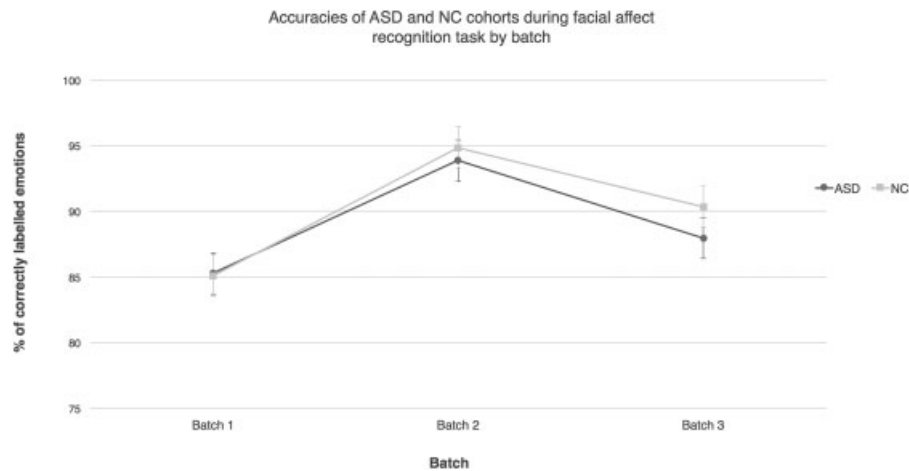
Fig. 3 Percentage of correct labeling during the computer task over the course of each batch.

By examining the confusion matrices before and after our Glass intervention (►Tables 3 and 4, respectively), we notice several trends. First, children were more likely to be NC if they correctly labeled neutral faces as "neutral." We noticed this correlation persisted before and after our Glass intervention (coefficients for Neutral–Neutral in batch 1: –6.93; batch 2: –0.94). Second, we observed that during batch 3, children who confused fear for surprised and vice versa were more likely to have ASD, and children who confused Fear with Disgust were more likely to be NC. The results seemed mixed during batch 1; however, the effect seemed larger in batch 3 (coefficient for Fear–Surprised: 4.76, Surprised–Fear: 13.16). Lastly, children were also more likely to have ASD if they confused Sad for Neutral, which remained consistent across batches (coefficient for Sad–Neutral in batch 1: 16.86, batch 3: 5.03).

Discussion

This feasibility study supported the hypothesis that Google Glass is a convenient wearable device to provide unobtrusive emotion-feedback cues to children with ASD. However, given the simplicity of the emotion recognition task and the average age of our study participants, it is not surprising to see little difference in the ability to detect emotions with and without emotion information, despite the previous research studies which have also found insignificant differences in

emotion recognition tasks between ASD participants and NC for this age range.[36,37] Further work will be necessary to assess the distinct differences between age-matched ASD and NC children. However, our confusion matrices provide further evidence to distinguish NC from children with ASD during facial affect recognition tasks.

Hypothesis 1

Our work confirmed that the Autism Glass learning aid fits comfortably and is not overstimulating for children with ASD. The Glass appropriately engages children with and without autism, providing feasibility of application for learning tasks including the emotion recognition learning attempted in our study. This finding supports the use of Google Glass for children with autism, expanding upon studies using portable devices and wearables to deliver behavioral intervention by bringing therapy to natural human interactions.[31,32]

Hypothesis 2

Supporting our second hypothesis, the emotion information provided by the Google Glass was associated with an increase in emotion labeling accuracy children during a facial affect recognition task. We see only a small drop in ASD scores relative to NC scores ($p = 0.310$) in the final batch and can hypothesize that this was due to fatigue during the task. The ASD cohort did not demonstrate significantly lower

Table 3 Coefficients of L_1-regularized logistic regression trained with batch 1 (before intervention)

TV/Response	Ha	An	Su	Ne	Fe	Sa	Di
Happy							
Angry		4.94					
Surprise					0.64		
Neutral				–6.93	0.21		
Fear			–2.16				
Sad		–4.50		16.86		–0.95	
Disgust							3.04

Table 4 Coefficients of L_1-regularized logistic regression trained with batch 3 (after intervention)

TV/Response	Ha	An	Su	Ne	Fe	Sa	Di
Happy							
Angry							2.65
Surprise					13.16		
Neutral				−0.94			
Fear		5.34	4.76				−12.47
Sad				5.03		0.30	−0.79
Disgust		−5.03			1.75		

accuracies during the facial affect recognition task in comparison to the NC cohort, however. This result supports research that states that children with ASD and NCs demonstrate no differences during facial affect recognition tasks of static images of basic emotions.[36,37] Future work is necessary to examine differences during nonstatic facial affect recognition tasks, and to examine differences in ability to recognize complex emotions.

Hypothesis 3

While the ASD and NC groups did not show statistically significant accuracy differences in their ability to label facial images with the correct emotion, they did show consistent differences in the emotions that they labeled incorrectly. Interestingly, the frequency of emotion confusion proved robust enough to enable statistically significant classification between the two groups, such that if children confused fear for surprise, surprise for fear, or sad for neutral, they would be classified as ASD.

We also note that algorithms trained to classify ASD versus NC given each participant's overall progression of accuracies (batch 1, batch 2, and batch 3), as well as algorithms trained on each participant's accuracies by expression, performed no better than chance using the same nested cross-validation scheme. This suggests that differences in expression recognition between ASD and NC exist but are subtle: the differences depend on which expressions were confused, rather than the overall accuracy of expression recognition.

Limitations

Our results and the ability to extrapolate meaning from these initial findings were limited by sample size, and the results mentioned in this discussion section require further replication. While we can claim distinct behavior between NC and ASD participants, due to a lack of control group that did not receive emotion data, we cannot conclude that the learning effect seen across batches was due to the device stimulus.

There are several limitations on the conclusions that we can draw from our study as a result of the participants included for analysis. For instance, while autism glass is a system intended to bring therapy to children at a younger age, we were only able to examine the feasibility of children ages 6 to 17 years of age in this study. Unfortunately, this age range is not representative of children who are in critical periods of development for cognition and speech, and therefore further feasibility testing on younger children is necessary.

In the present study, we observed a gender ratio of 13 males to 3 females in our ASD cohort. Though males are substantially more likely to be diagnosed with autism than females by an average ratio of 4 males:1 female,[61,62] the reported imbalance still presents a gender bias between our sample population of children with ASD to our NC children.

Furthermore, 3 of our 16 ASD participants had an ABIQ lower than 80 (between 55 and 79), suggesting that, based on ABIQ scores, 13 of our recruited ASD participants can be classified as children with high-functioning ASD. This limits our findings for Hypotheses 2 and 3 to generalize to the greater ASD spectrum, as their performance on the facial affect recognition task was likely more accurate than most children on the ASD spectrum.

Additionally, the six ASD participants who were not able to complete study procedures (mentioned in the "Participant Demographics" section above) were not included in analysis for Hypotheses 2 and 3. Although they wore the glass with ease for their study visit, which supports Hypothesis 1, they experienced technological failures with the software and their data were discarded. As described in the "Participant Demographics" section, these children were age-, gender-, and intelligence quotient (IQ)-matched with the rest of our cohort. We were able to include their results to support Hypothesis 1; however, we had to discard their data from analyses for Hypotheses 2 and 3. These children did not represent a low functioning participants, nor were we able to recruit a representative population of low-functioning participants, likely skewing our results for Hypotheses 2 and 3 toward a null finding, further supporting that our results may misrepresent the full autism spectrum population.

The first four study participants, all NC males between 7 and 11 years old with ABIQs ranging from 85 to 106, received only visual data on the heads-up display in the form of the word corresponding to the correct expression. Through qualitative feedback that these participants provided to the study team, and after a preliminary analysis, we determined that three participants were unable to read the displayed words. The three participants who were unable to read the

visual cues scored in the 30th, 58th, and 61st percentiles (with 95% confidence interval [CI]) on the verbal component of the ABIQ. We did not assess the ability of the ASD participant population in reading the visual cues provided by the glasses, nor did we display the visual cues for longer time periods to then assess participants' ability to read the cues.

Lastly, our Glass prototype in this study did not examine how receiving emotion data from this device during facial affect recognition tasks will generalize to more meaningful social interactions. We address this limitation below.

Conclusion

The results from this feasibility and pilot study support literature that state children with ASD have no less accuracy than NC in a controlled, image-based facial affect recognition task; however, our results also support findings that children with ASD more often confuse fear for surprise (and vice versa), and confuse sad for neutral more often than their NC

peers.[34,36,38,39,63] Furthermore, if children correctly labeled neutral faces, they were more likely to be NC, supporting previous research findings that children with ASD have difficulties labeling neutral faces correctly.[34,40]

Future Directions

The delivery of the behavioral intervention programs in today's health care system is increasingly bottlenecked because the number of behavioral therapists is far outstripped by the number of children in need of care.[1–5,64] Motivated by the need for more scalable care, we believe that a lightweight and mobile therapeutic tool may prove efficacious for augmenting therapy needed by children with autism. This mobile-therapeutic tool combines live, natural interactions with evidence-based properties that form the foundation of teaching strategies and behavioral therapies as a whole, such as positive reinforcement, iterative training, and structured feedback. Focusing on deficits associated with facial affect recognition, our proposed therapeutic learning tool is an artificial
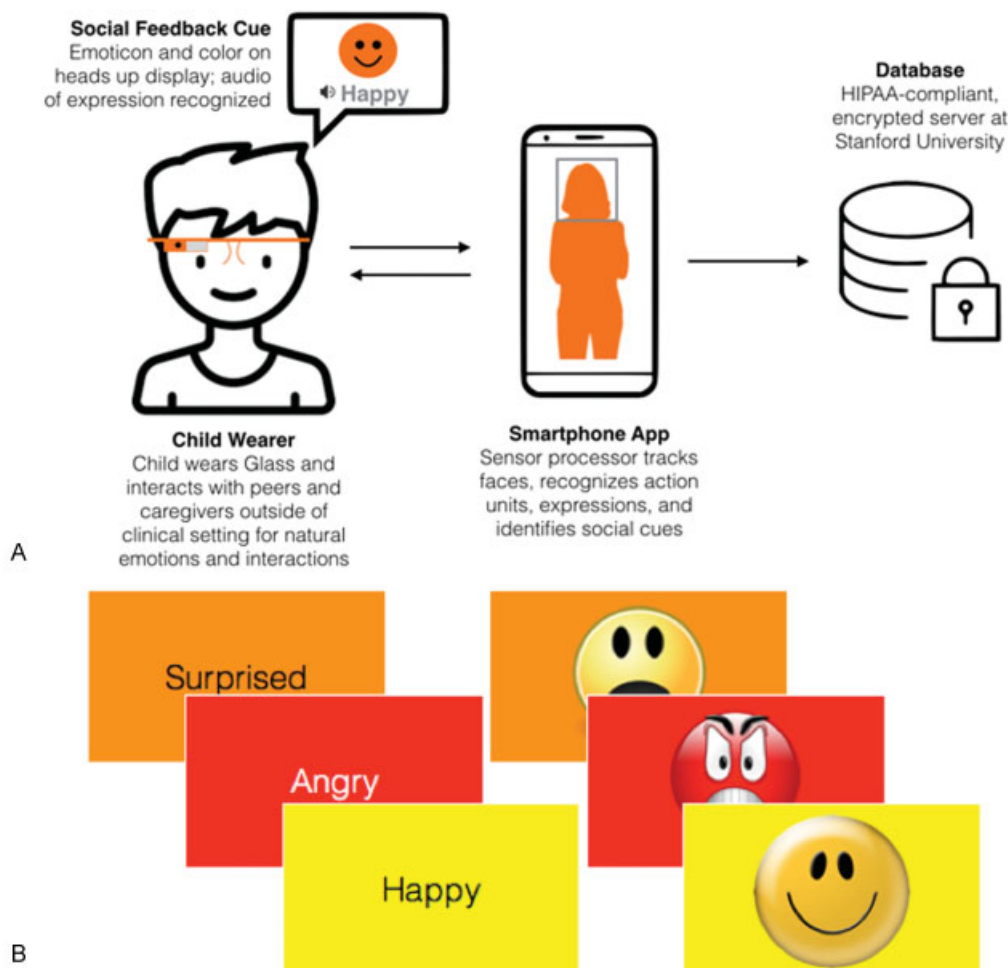

Fig. 4 Feedback loop and system architecture overview for the Superpower Glass system. (**A**) The system overview for real-time facial expression recognition. This shows how the device can communicate with a smartphone to enable feedback choice directly from a control center within the App. (**B**) It depicts the potential feedback options from which a user of this wearable therapy can choose. This includes audio feedback, visual feedback (emojis, words, colors), and a combination of audio and visual feedback. Visual feedback cues are provided to the child via the Glass' heads-up display, whereas audio cues are delivered using the bone conducting speaker within the Glass units themselves.

intelligence system for automatic facial expression recognition that labels facial expressions in real time via Google Glass, recognizing emotions in the faces of conversation partners and delivering real-time emotion cues to the wearer, a child with autism. The system uses the Glass's outward-facing camera to read facial expressions and an emotion classifier to compute and return to the child wearer the emotion of the person in the field of view as real-time social emotion cues as audio and/or visual information provided by the Glass unit (►**Fig. 4**). Visual representations can facilitate an increased rate of learning in children with autism, who often learn best via visual learning, and that pairing these visual representations with verbal cues can maximize impact on learning.[25,26,30] Our proposed therapeutic tool can supplement existing therapeutic approaches and may be used either by trained behavioral interventionists or independently by children with autism and their caregivers.

Validation through a Randomized Control Trial

Now that we have confirmed that children with ASD respond positively to wearing the Glass and to receiving visual/audio emotion data from it, we have created a working prototype system. Our prototype device consists of a Google Glass paired with a phone (►**Fig. 4**), upon which we have optimized a system of real-time facial expression recognition and delivery of accurate emotion information. We intend to test the prototype in a randomized control trial to determine under what circumstances this tool can be used by children and family members outside of clinical settings, and how this tool might help alleviate gaps in care due to lack of access (e.g., while on waiting lists or for those in remote and underserved areas). We will evaluate the prototype's ability to teach children with ASD how to interpret emotion in faces, to improve overall social awareness, to decrease social anxiety, and to increase eye contact during social interactions.

Clinical Relevance Statement

This study supports the feasibility and use of mobile wearable tools as a learning aid to children with ASD. This pilot study enables us to move forward with creating a mobile at-home behavioral tool, Superpower Glass, which can have many implications for clinical implementation.

Multiple Choice Questions

1. During which batch was the Autism Glass intervention introduced to the children in this study?
   a. Batch 1
   b. Batch 2
   c. Batch 3
   d. None of the batches involved an intervention.

   Correct answer: The correct answer is b, batch 2. Batch 1 assessed the children's baseline during the task. The intervention was introduced for Batch 2 as an aid while completing the same task in Batch 1. Batch 3 removed the aid to see if on their own, they were able to label emotions more accurately.

2. How did children with ASD compare with their neurotypical controls during the emotion recognition task?
   a. Children with ASD performed with statistically significant higher accuracy than neurotypical controls.
   b. Children with ASD performed with statistically significant lower accuracy than neurotypical controls.
   c. Children with ASD were not able to recognize "Happy," unlike the neurotypical controls.
   d. Children with ASD and neurotypical controls did not differ in accuracy by any statistically significant result.

Correct Answer: The correct answer is d. Children with ASD and neurotypical controls had very similar accuracy results. However, children with ASD differed from neurotypical controls in that they more often confused certain emotions, and had different confusion models than neurotypical controls. (See the "Results" section for more specifics.)

References

1 Buescher AV, Cidav Z, Knapp M, Mandell DS. Costs of autism spectrum disorders in the United Kingdom and the United States. JAMA Pediatr 2014;168(08):721–728

2 Christensen DL, Baio J, Van Naarden Braun K, et al; Centers for Disease Control and Prevention (CDC). Prevalence and characteristics of autism spectrum disorder among children aged 8 years–Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2012. MMWR Surveill Summ 2016; 65(03):1–23

3 Christensen DL, Bilder DA, Zahorodny W, et al. Prevalence and characteristics of autism spectrum disorder among 4-year-old children in the Autism and Developmental Disabilities Monitoring Network. J Dev Behav Pediatr 2016;37(01):1–8

4 Nicholas JS, Charles JM, Carpenter LA, King LB, Jenner W, Spratt EG. Prevalence and characteristics of children with autism-spectrum disorders. Ann Epidemiol 2008;18(02):130–136

5 Robins DL. Prevalence counts: commentary on "prevalence and characteristics of autism spectrum disorder among 4-year-old children in the Autism and Developmental Disabilities Monitoring Network.". J Dev Behav Pediatr 2016;37(01):80–82

6 Dawson G, Rogers S, Munson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. Pediatrics 2010;125(01):e17–e23

7 Howlin P, Goode S, Hutton J, Rutter M. Adult outcome for children with autism. J Child Psychol Psychiatry 2004;45(02):212–229

8 Landa RJ, Holman KC, Garrett-Mayer E. Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. Arch Gen Psychiatry 2007;64(07):853–864

9 Landa RJ, Holman KC, O'Neill AH, Stuart EA. Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: a randomized controlled trial. J Child Psychol Psychiatry 2011;52(01):13–21

10 Palumbo L, Burnett HG, Jellema T. Atypical emotional anticipation in high-functioning autism. Mol Autism 2015;6:47

11 Sasson NJ, Pinkham AE, Weittenhiller LP, Faso DJ, Simpson C. Context effects on facial affect recognition in schizophrenia and autism: behavioral and eye-tracking evidence. Schizophr Bull 2016;42(03):675–683

12 Xavier J, Vignaud V, Ruggiero R, Bodeau N, Cohen D, Chaby L. A multidimensional approach to the study of emotion recognition in autism spectrum disorders. Front Psychol 2015;6:1954

13 Marsh AA, Blair RJR. Deficits in facial affect recognition among antisocial populations: a meta-analysis. Neurosci Biobehav Rev 2008;32(03):454–465

14 Blair RJR. Responding to the emotions of others: dissociating forms of empathy through the study of typical and psychiatric populations. Conscious Cogn 2005;14(04):698–718

15 Rump KM, Giovannelli JL, Minshew NJ, Strauss MS. The development of emotion recognition in individuals with autism. Child Dev 2009;80(05):1434–1447

16 Yirmiya N, Sigman MD, Kasari C, Mundy P. Empathy and cognition in high-functioning children with autism. Child Dev 1992;63(01):150–160

17 Burnette CP, Mundy PC, Meyer JA, Sutton SK, Vaughan AE, Charak D. Weak central coherence and its relations to theory of mind and anxiety in autism. J Autism Dev Disord 2005;35(01):63–73

18 Schreibman L, Dawson G, Stahmer AC, et al. Naturalistic developmental behavioral interventions: empirically validated treatments for autism spectrum disorder. J Autism Dev Disord 2015;45(08):2411–2428

19 Nash JM. Fertile minds. Time 1997;149(05):48–56

20 Phillips DA, Shonkoff JP. From Neurons to Neighborhoods: The Science of Early Childhood Development. Sydney, Australia: National Academies Press; 2000

21 Dawson G, Bernier R. A quarter century of progress on the early detection and treatment of autism spectrum disorder. Dev Psychopathol 2013;25(4 Pt 2):1455–1472

22 Dawson G, Burner K. Behavioral interventions in children and adolescents with autism spectrum disorder: a review of recent findings. Curr Opin Pediatr 2011;23(06):616–620

23 Dawson G, Webb SJ, McPartland J. Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. Dev Neuropsychol 2005;27(03):403–424

24 Chen SH, Bernard-Opitz V. Comparison of personal and computer-assisted instruction for children with autism. Ment Retard 1993;31(06):368–376

25 Golan O, Ashwin E, Granader Y, et al. Enhancing emotion recognition in children with autism spectrum conditions: an intervention using animated vehicles with real emotional faces. J Autism Dev Disord 2010;40(03):269–279

26 Grynszpan O, Weiss PL, Perez-Diaz F, Gal E. Innovative technology-based interventions for autism spectrum disorders: a meta-analysis. Autism 2014;18(04):346–361

27 Mitchell P, Parsons S, Leonard A. Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. J Autism Dev Disord 2007;37(03):589–600

28 Moore M, Calvert S. Brief report: vocabulary acquisition for children with autism: teacher or computer instruction. J Autism Dev Disord 2000;30(04):359–362

29 Parsons S, Mitchell P, Leonard A. The use and understanding of virtual environments by adolescents with autistic spectrum disorders. J Autism Dev Disord 2004;34(04):449–466

30 Pennington RC. Computer-assisted instruction for teaching academic skills to students with autism spectrum disorders: a review of literature. Focus Autism Other Dev Disabl 2010;25(04):239–248

31 Madsen M, El Kaliouby R, Goodwin M, Picard R, eds. 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

32 Liu R, Salisbury JP, Vahabzadeh A, Sahin NT. Feasibility of an autism-focused augmented reality smartglasses system for social communication and behavioral coaching. Front Pediatr 2017;5:145

33 Washington P, Voss C, Haber N, et al, eds. A wearable social interaction aid for children with autism. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM; 2016

34 Harms MB, Martin A, Wallace GL. Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. Neuropsychol Rev 2010;20(03):290–322

35 Maglione MA, Gans D, Das L, Timbie J, Kasari C; Technical Expert Panel; HRSA Autism Intervention Research – Behavioral (AIR-B) Network. Nonmedical interventions for children with ASD: recommended guidelines and further research needs. Pediatrics 2012;130(Suppl 2):S169–S178

36 Castelli F. Understanding emotions from standardized facial expressions in autism and normal development. Autism 2005;9(04):428–449

37 Ozonoff S, Pennington BF, Rogers SJ. Are there emotion perception deficits in young autistic children? J Child Psychol Psychiatry 1990;31(03):343–361

38 Baron-Cohen S, Golan O, Ashwin E. Can emotion recognition be taught to children with autism spectrum conditions? Philos Trans R Soc Lond B Biol Sci 2009;364(1535):3567–3574

39 Corden B, Chilvers R, Skuse D. Avoidance of emotionally arousing stimuli predicts social-perceptual impairment in Asperger's syndrome. Neuropsychologia 2008;46(01):137–147

40 Berggren S, Engström AC, Bölte S. Facial affect recognition in autism, ADHD and typical development. Cogn Neuropsychiatry 2016;21(03):213–227

41 Fridenson-Hayo S, Berggren S, Lassalle A, et al. Basic and complex emotion recognition in children with autism: cross-cultural findings. Mol Autism 2016;7(01):52

42 Hopkins IM, Gower MW, Perez TA, et al. Avatar assistant: improving social skills in students with an ASD through a computer-based intervention. J Autism Dev Disord 2011;41(11):1543–1555

43 Kuusikko S, Haapsamo H, Jansson-Verkasalo E, et al. Emotion recognition in children and adolescents with autism spectrum disorders. J Autism Dev Disord 2009;39(06):938–945

44 Russo-Ponsaran NM, Evans-Smith B, Johnson J, Russo J, McKown C. Efficacy of a facial emotion training program for children and adolescents with autism spectrum disorders. J Nonverbal Behav 2016;40(01):13

45 Ben-Sasson A, Hen L, Fluss R, Cermak SA, Engel-Yeger B, Gal E. A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. J Autism Dev Disord 2009;39(01):1–11

46 Hilton CL, Harper JD, Kueker RH, et al. Sensory responsiveness as a predictor of social severity in children with high functioning autism spectrum disorders. J Autism Dev Disord 2010;40(08): 937–945

47 Leekam SR, Nieto C, Libby SJ, Wing L, Gould J. Describing the sensory abnormalities of children and adults with autism. J Autism Dev Disord 2007;37(05):894–910

48 Liss M, Saulnier C, Fein D, Kinsbourne M. Sensory and attention abnormalities in autistic spectrum disorders. Autism 2006;10 (02):155–172

49 Tonacci A, Billeci L, Tartarisco G, et al. [Formula: see text]Olfaction in autism spectrum disorders: a systematic review. Child Neuropsychol 2017;23(01):1–25

50 Allen CW, Silove N, Williams K, Hutchins P. Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. J Autism Dev Disord 2007;37(07):1272–1278

51 Association WM. World Medical Association Declaration of Helsinki: recommendations guiding physicians in biomedical research involving human subjects [Internet]. Somerset West, RSA: 48th World Medical Association General Assembly; 1996 Oct. [cited 2010 Jun 8]Nord Med 1992;107:24–25

52 Roid GH. Stanford-Binet Intelligence Scales. Itasca, IL: Riverside Publishing; 2003

53 Constantino JN, Gruber CP. Social Responsiveness Scale. 2nd ed (SRS-2). Los Angeles, CA: Western Psychological Services; 2012

54 LoBue V, Thrasher C. The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. Front Psychol 2015;5:1532

55 Ekman P, Friesen WV, O'Sullivan M, et al. Universals and cultural differences in the judgments of facial expressions of emotion. J Pers Soc Psychol 1987;53(04):712–717

56 Dichter GS, Felder JN, Green SR, Rittenberg AM, Sasson NJ, Bodfish JW. Reward circuitry function in autism spectrum disorders. Soc Cogn Affect Neurosci 2012;7(02):160–172

57 DePape A-MR, Hall GB, Tillmann B, Trainor LJ. Auditory processing in high-functioning adolescents with Autism Spectrum Disorder. PLoS One 2012;7(09):e44084

58 Smith EG, Bennetto L. Audiovisual speech integration and lipreading in autism. J Child Psychol Psychiatry 2007;48(08):813–821

59 Murphy KP. Machine Learning: A Probabilistic Perspective. Cambridge, MA: MIT Press; 2012

60 Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 2002;15(01):1–25

61 Constantino JN, Charman T. Gender bias, female resilience, and the sex ratio in autism. J Am Acad Child Adolesc Psychiatry 2012; 51(08):756–758

62 Dworzynski K, Ronald A, Bolton P, Happé F. How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? J Am Acad Child Adolesc Psychiatry 2012;51 (08):788–797

63 Howard MA, Cowell PE, Boucher J, et al. Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. Neuroreport 2000;11(13):2931–2935

64 Laski KE, Charlop MH, Schreibman L. Training parents to use the natural language paradigm to increase their autistic children's speech. J Appl Behav Anal 1988;21(04):391–400