Technology for Just-In-Time *In-Situ* Learning of Facial Affect for Persons Diagnosed with an Autism Spectrum Disorder

Miriam Madsen
MIT Media Lab
20 Ames Street
Cambridge MA 02139
+001-617-894-0101
mish@mit.edu

Rana el Kaliouby
MIT Media Lab
20 Ames Street
Cambridge MA 02139
+001-617-253-8628
kaliouby@media.mit.edu

Matthew Goodwin
Groden Center
86 Mount Hope Ave.
Providence RI 02906
+001-401-274-6310
mgoodwin@grodencenter.org

Rosalind Picard
MIT Media Lab
20 Ames Street
Cambridge MA 02139
+001-617-253-0611
picard@media.mit.edu

ABSTRACT

Many first-hand accounts from individuals diagnosed with autism spectrum disorders (ASD) highlight the challenges inherent in processing high-speed, complex, and unpredictable social information such as facial expressions in real-time. In this paper, we describe a new technology aimed at helping people capture, analyze, and reflect on a set of social-emotional signals communicated by facial and head movements in live social interaction that occurs with their everyday social companions. We describe our development of a new combination of hardware using a miniature camera connected to an ultramobile PC together with custom software developed to track, capture, interpret, and intuitively present various interpretations of the facial-head movements (e.g., presenting that there is a high probability the person looks "confused"). This paper describes this new technology together with the results of a series of pilot studies conducted with adolescents diagnosed with ASD who used the technology in their peer-group setting and contributed to its development via their feedback.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and behavioral sciences – *psychology.*

General Terms

Design, Experimentation, Human Factors.

Keywords

Autism spectrum disorders, Asperger's Syndrome, Facial Affect, Facial Expressions, Affective computing

ASSETS'08, October 13–15, 2008, Halifax, Nova Scotia, Canada.

1. INTRODUCTION

Autism spectrum disorder (ASD) encompasses a wide range of developmental issues that typically manifest in childhood, with mild to dramatic implications for social interaction, communication abilities, sensory perception, and repetitive behaviors and interests [1]. ASD includes Autistic Disorder (AD); Asperger syndrome (AS); and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS).

Figure 1: Autistic adolescents using the wearable camera system and facial analysis software to capture and analyze their social interactions.

AD, also known as "classical autism," affects behavior in the areas of social interaction, verbal and nonverbal communication, and exhibited interests. Some frequently associated symptoms seen in autistic persons[1] include discomfort with eye contact; low levels of expressed empathy; diminished or nonexistent verbal

[1] Many people diagnosed as having an ASD have expressed a preference for the usage of "autistic person" instead of "person with autism." See Sinclair 1999 http://web.syr.edu/~jisincla/person_first.htm for an essay on this respectful use.

communication; increased sensitivity to touch; repetitive movements, such as flapping hands or rocking; late onset of speech; and difficulty in perceiving and reacting to others' conversational and emotional cues [2].

AS, sometimes also called "High-Functioning Autism" (HFA), is primarily characterized by significant preoccupation with a particular subject, repetitive routines, lack of coordination, and inappropriate behavior in social situations. Speech patterns can also be directly affected in this population; for instance, speech may be overly formal or monotonal, and a person may exhibit a lessened capacity to understand figurative speech. People diagnosed with AS can usually function in normal society but frequently experience difficulties relating to nonstandard communication behaviors and problems with social interaction [3].

The PDD-NOS diagnosis encompasses a series of symptoms frequently found both in people diagnosed with autism and people with developmental delay and cognitive impairments. PDD-NOS is essentially a 'subthreshold' condition in which an insufficient number of the standard criteria for autism are met; however, PDD-NOS itself is a clinical diagnosis reflecting pervasive social and/or cognitive difficulties [4].

Our goal in this paper is to illustrate some of the methods we have tried to use for individuals diagnosed with ASD to help them with social communication. Individuals in these situations often struggle with multiple levels of social interactions, due to inherent difficulties picking up on interpersonal cues that result in nonstandard socialization. We propose that an ideal social skills intervention would have a number of components that complement a technical solution to the problem at hand.

(1) It should be personally situated, allowing the person to explore and learn in a social environment that is as realistic and natural as possible. This may provide intrinsic motivation to learn since the training would be focused on interacting with people whom they know and care about. It may also be better for generalizing the knowledge he or she has acquired in familiar settings with familiar persons to other less familiar situations and interaction partners.

(2) Just-in-time training may enable a person to get feedback at the moment of need, when learning is most sought. Social situations can be overwhelming, so another key component of our approach is to use technology that enables systematic exploration of social interactions, thereby helping users gain incremental insights into their expressions and those of others.

(3) Easy-to-use and fun technologies should allow more opportunities for learning than is typically available in infrequent, school-based therapeutic sessions.

Thus, our research has been oriented not only around helping these individuals interact with others, but also helping them thoroughly understand their own patterns of interaction. The principal contribution of this work is a novel paradigm oriented around *in situ* social learning technology. Our system is framed around live computer-assisted facial affect analysis dedicated to helping the recipients of the intervention develop strategies for social interaction that can be generalized to other situations. By creating non-invasive and simple-to-use technology, we hope to make our system accessible to a broad range of people with autism.

The paper is organized as follows: the following section highlights social interaction difficulties in autism and surveys existing technologies for teaching facial expression understanding to individuals on the spectrum. Section 3 describes the mobile PC platform that we use and presents the facial analysis and results visualization software that we developed. Sections 4 and 5 describe the pilot experiments and present the results and lessons learned, and Section 6 concludes the paper and outlines several future directions of this work.

2. SOCIAL DIFFICULTIES IN AUTISM SPECTRUM DISORDER

There are many socialization difficulties associated with a diagnosis of ASD. Individuals face major issues in effective communication stemming from factors as diverse as stress incurred through eye contact; difficulty of timing in oral exchanges; non-standard analysis of conversation dynamics, often leading to the missing of standard conversational cues; and particular issues with conveying and interpreting emotion in speech in a normative fashion. At the same time, many people on the autism spectrum have strengths related to systemizing information [14], sometimes accompanied by enhanced perceptual function [15]. We have aimed in this work to address autistic social difficulties and leverage autistic processing strengths, by providing a forum in which autistic individuals can analyze their interactions in two separate ways: both in real-time, so that a quantitative analysis of facial expressions (assisted by an affect-recognizing program and live video input) can be observed by the child and a teacher; and in a recorded setting, in which the child can record naturally-evoked facial expressions and use the facial-expression reading program to analyze these expressions later. Through this dual method, we hope to encourage a critical-thinking experience of facial affect interpretation that will come more naturally to individuals on the spectrum, as they tend to be more adept at managing quantitatively presented information than they are at processing qualitative and highly subjective information such as interpersonal social cues.

A number of intervention approaches have been developed to teach people on the autism spectrum social reading skills. A majority of these are structured programs that either focus on prototypic basic emotions or on social skills training such as reducing socially inappropriate behavior and improving personal hygiene [5, 6]. Structured programs that teach the basic emotions use carefully constructed facial schematics [7, 8] and other nonverbal stimuli [9]. These teaching approaches are generally accepted by individuals diagnosed with HFA/AS, but there is little scientific evidence that acquired skills generalize to real-time, real-world interactions. Mind Reading DVD [10] is an interactive DVD guide to emotions with video clips that portray 421 mental state concepts. To the best of our knowledge, it is the only available system for teaching autistic people facial expressions and mental states beyond a simple set of basic emotions. It presents interactive games for players to practice their ability to recognize a wide range of emotions and social signals from face-videos.

Our approach with wearable technologies offers, for the first time, the ability to conduct *just-in-time in situ* assistance to help individuals with HFA/AS to learn facial expressions and underlying emotions in their own specific natural environments. It achieves this through a mixture of new hardware and new software capabilities.

3. MATERIALS AND METHODS

3.1 Overview

The principal physical components of our system include a Samsung ultramobile computer and attached Logitech camera, as detailed in section 3.2. In sections 3.3 and 3.4, we detail the two custom software applications that are used on the computer to analyze and interact with the facial affect information.

3.2 Hardware

Our goal is to help autistic individuals interact with people in a live setting and develop a qualitative intuition of the visible facial communication of their own and others' social-emotional states. For example, we want them to see, in real time, if somebody is looking confused or disagreeing with what they are saying, or if somebody is looking interested or smiling. To that end, we wanted to allow them to process – both live and in later recorded contexts – their facial expressions and the expressions of their friends. We chose to use Samsung ultramobile PCs (8.96" x 4.88" x 0.93" in size, 1.52 lbs) to process the live recorded video. We used a very small Logitech USB camera attached to the ultramobile computers to assist with mobile camera processing. The hardware is shown in Figures 2 and 3.

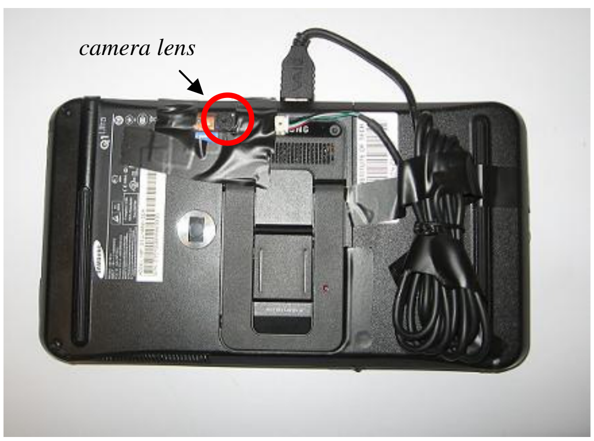

Figure 2: back of Samsung mini-computer with attached camera for video capture

Figure 3: front of Samsung mini-computer.

3.3 Facial Analysis System

To interpret the live video feeds, we use the facial analysis framework and C++ software of el Kaliouby [11]. This framework combines bottom-up vision-based processing of the face and head movements (e.g., a head nod or smile) with top-down predictions of mental state models (e.g., interest and agreeing) to interpret the meaning underlying head and facial signals over time. A multilevel probabilistic algorithm using Dynamic Bayesian Networks models the hierarchical way with which people are thought to perceive facial and other human behavior [12] and handles the uncertainty inherent in the process of attributing mental states to others.

At camera speed, the inference system locates and tracks 24 feature points on the face and uses motion, shape and color deformations of these features to identify 20 facial and head movements (e.g., head pitch, lip corner pull) or action units from the Facial Action Coding System [13] and 11 communicative gestures (e.g., head nod, smile, eyebrow flash). The person's affective-cognitive state is then inferred based on what the model has learned from training examples that relate combinations of these facial and head movements to higher-level "emotion" labels. Our inference approach here is unique in that it supports the inference of six states of mind that extend beyond the basic emotions. These six states are: *agreeing, concentrating, disagreeing, interested, thinking* and *confused*. The first three of these states (which in this paper and in the work with the kids we loosely refer to as "emotions") are designated in our system as "positive" emotions, since they frequently indicate that a conversation with a person on the spectrum is proceeding productively; the last three, including "Thinking", may indicate that a conversation has reached a point where the speaker may want to reiterate or rephrase a previous point in order to make sure the listener understands. The other novelty component of the system is that it accounts for the richness and complexity inherent in facial affect signals by allowing for multiple mappings from a face (e.g., a smile) to a mental state (e.g., *agreeing, interested*).

3.4 Emotion Bubbles

In preparation for this project, we designed the Emotion Bubbles interface for the Facial Analysis System using Processing, a Graphical User Interface-oriented language with a JAVA base. In replacing the default output of the Facial Analysis system, which had consisted of a line graph on which all six mental states were tracked over time, the Emotion Bubbles interface was specifically designed to provide information about levels of emotions in a simple and intuitive way that would be easily accessible to individuals on the spectrum. The numeric level of each emotion was therefore represented as the percentage of a bubble's full radius displayed on screen. Each emotion displayed was assigned a particular color. As given above, the "positive" emotions were assigned "cool" colors (green, blue, and purple) indicating a productive state, and the "negative emotions" were assigned "warm" colors (red, orange, and yellow) indicating that the user of the interface should be aware of a possible conversational impediment. Each bubble is updated dynamically with every new output line from the Facial Analysis System. The output frequency of the Facial Analysis System's text-based output is once per every five frames. Given our particular hardware for this project (the Logitech mini-camera with an input frame rate of about 5 frames per second) this corresponds to the Emotion

Bubbles interface updating two to three times per second. This update rate was frequent enough for our needs in terms of giving the adolescents fairly frequent information about the live or recorded video they were analyzing.

The Facial Analysis System and the Emotion Bubbles interface were presented side-by-side on the Samsung screen, as shown in Figures 4 and 5. The Facial Analysis System window was retained to present information about which parts of the face were being tracked; the Emotion Bubbles front end was added to provide an intuitive view of which emotions were currently being understood from the given input by the Facial Analysis System.

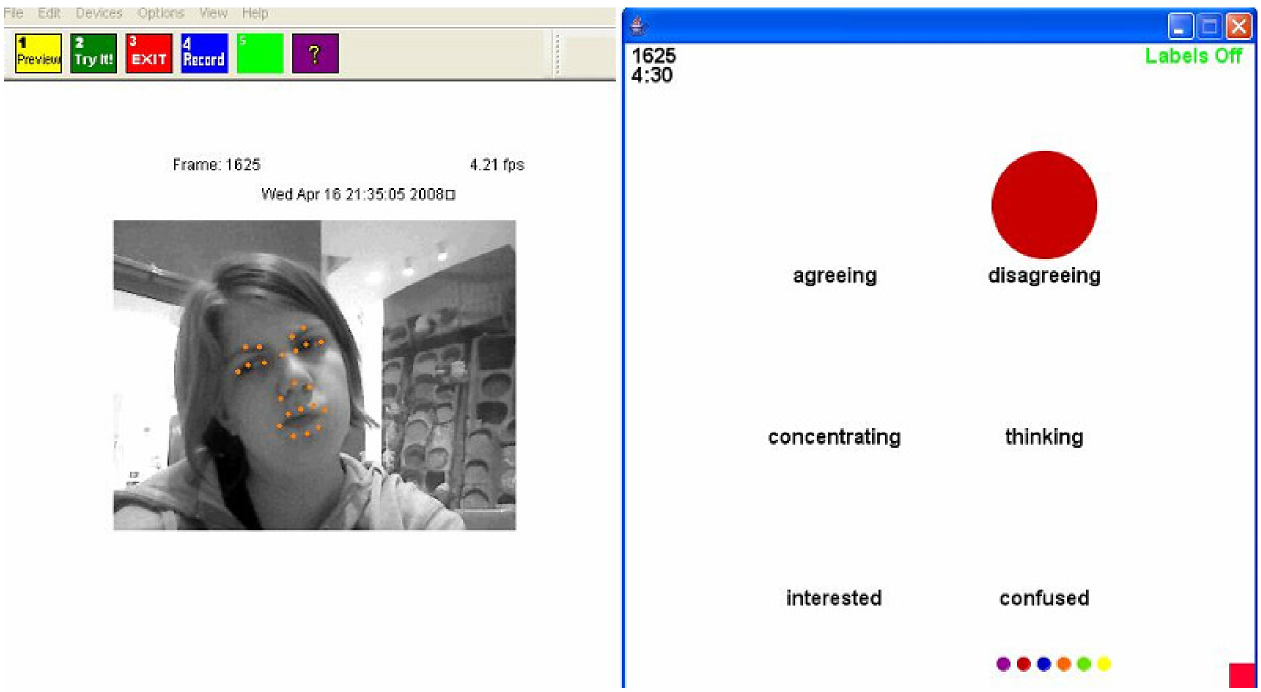

Figure 4: Illustration of Facial Analysis System video input and Emotion Bubbles interface.

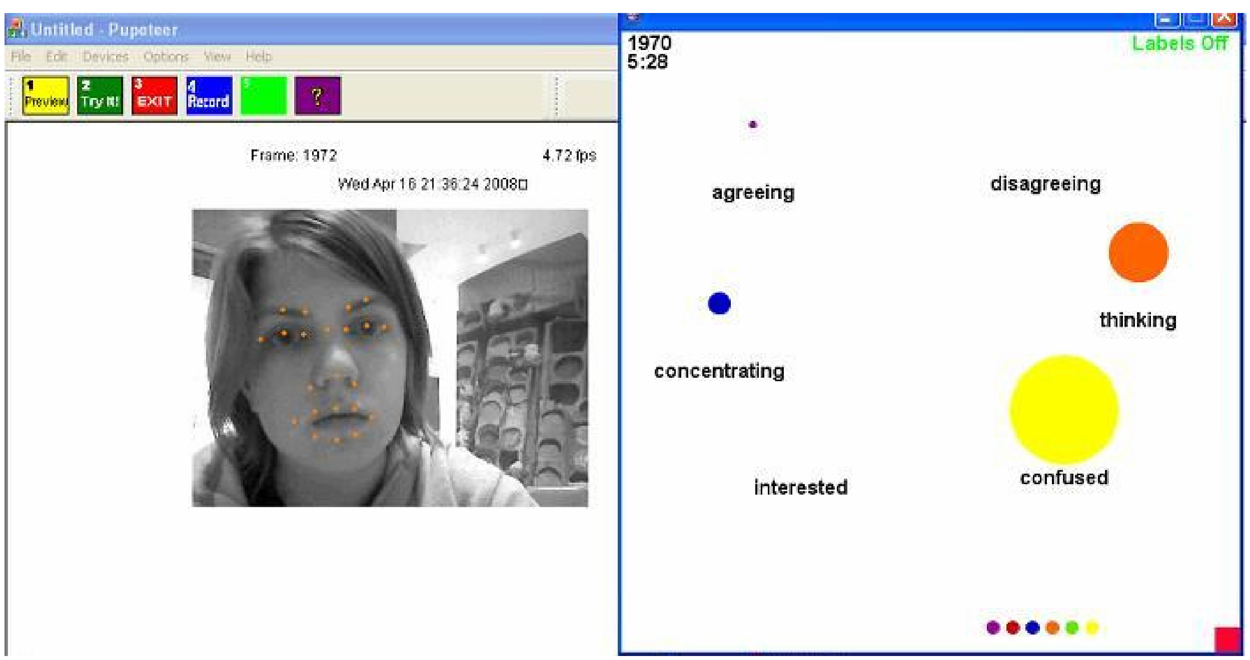

Figure 5: Illustration of Facial Analysis System video input and Emotion Bubbles interface.

4. PILOT EXPERIMENT

4.1 General information

We piloted the Facial Analysis System/Emotion Bubbles software combination on the Samsung/Logitech hardware at the Groden Center in Providence, RI, a facility that provides services to individuals with diagnoses of ASD as well as individuals with other developmental disabilities.

We visited the Groden Center multiple times to run our pilot; between visits, we iterated our protocol and software. With the help of the Groden Center staff, we ran two different types of pilots; one in which the individuals worked with the software in an essentially unstructured way, and another in which they recorded video to try to elicit particular expressions from their friends and then reviewed the video with the Facial Analysis System and the Emotion Bubbles later to gauge their own success.

Both pilots were run during all visits with the same three adolescent boys, diagnosed with AD, AS, or PDD-NOS.

4.2 Training

Prior to introducing the Facial Analysis System/Emotion Bubbles combination in a group setting, we showed each child in the pilot the interface in a separate setting. Using a standard laptop computer with a front-facing camera with which the child could watch as the Facial Analysis System analyzed their own facial expressions, we verbally assisted each child in understanding the significance of the bubbles and their relationship to the facial expressions displayed in the Facial Analysis System. We then conducted a series of informal exercises – asking, for example, whether they agreed with the statement “Macs are cool” for Windows aficionados and then pointing out that the Disagreeing bubble became larger as Facial Analysis System analyzed their facial expression.

4.3 Live video analysis

The live video analysis task essentially relied on leading the individuals to their own understanding of the ability to quantitatively analyze facial expressions in a group setting, thus developing their qualitative intuition about affective communication. Depending on the cognitive ability and verbal level of the participants watching and being analyzed by the Facial Analysis System, the experimenters would hold the computer, manage the interaction verbally, or neither, the latter condition becoming more frequent as the participants became more comfortable with interacting with the Facial Analysis System and the Emotion Bubbles. In situations where the experimenters did direct the interaction, our input was generally designed to help the participants elicit responses with specific facial expressions from their friends, such as prompting the child with the computer to ask the child being viewed on the Facial Analysis System screen, “Do you like the Red Sox?” with the goal of causing the Agreeing bubble to grow in size. Figure 1 illustrates how the participants actively used the technology while communicating with one another; Figure 6 demonstrates how a staffer was able to help a participant hold the computer for greater accuracy in facial expression analysis.

4.4 Recorded video analysis

During our second visit, we ran a task to help the individuals understand how they might try to elicit particular responses in a given social context. We asked them to record a series of facial expressions with the general goal of getting faces that would elicit the “agreeing” emotion bubble. We verbally phrased the task to emphasize the importance of getting the emotion in a natural conversation with the correct Emotion Bubbles interface response, i.e., “Can you get him to make some faces that look like he’s *agreeing* with you? It’s like trying to get the purple bubble to get bigger on the computer.”

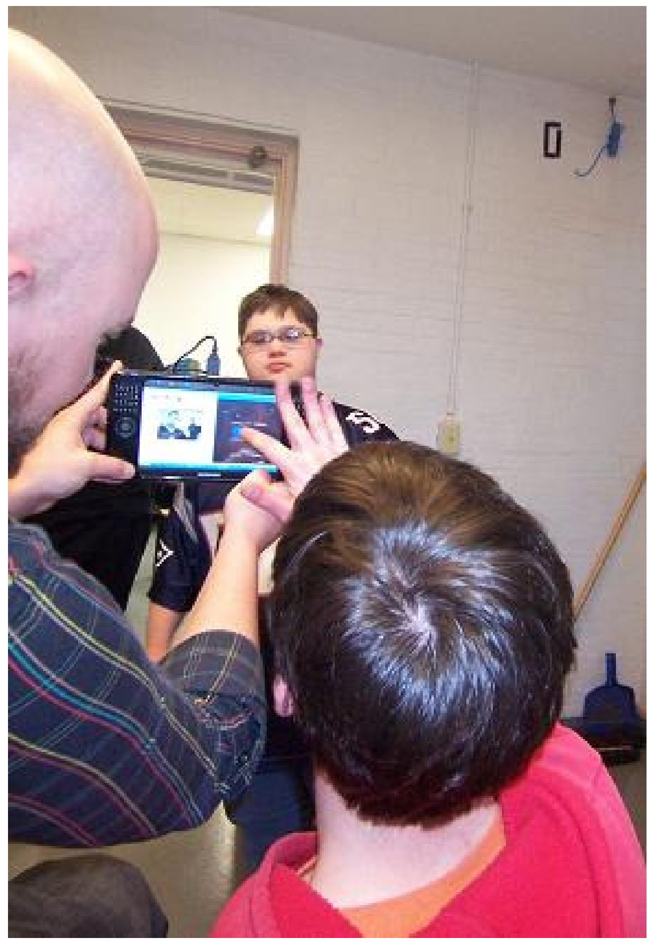

Figure 6: Groden Center staff assisting student with Facial Analysis System/Emotion Bubbles video analysis.

After asking each child to record a few minutes of their friends making expressive faces with the goal of getting faces that would register as ‘agreeing’, we separately took each child to a different room to review their video and talk about the success of their methods in eliciting the “correct” expression. This one-on-one analysis with the child was also an opportunity to talk about how different expressions produced certain bubbles and bubble combinations. Figure 7 shows an experimenter reviewing a participant's recorded video and discussing the Emotion Bubble results with the participant.

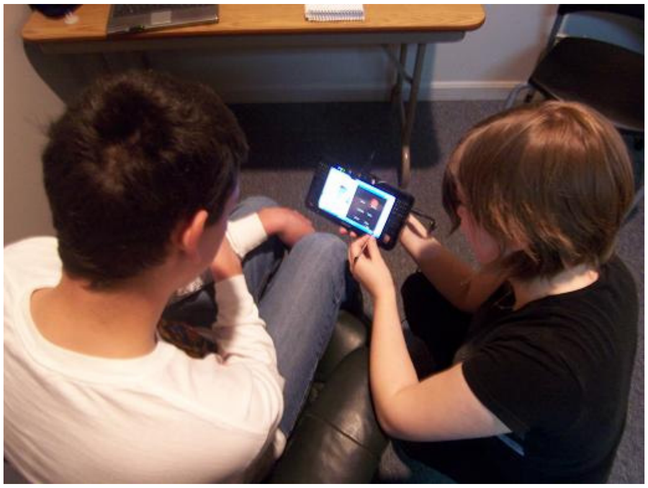

Figure 7: Student reviews recorded video with experimenter using the Facial Analysis System and Emotion Bubbles.

5. RESULTS

The goal of these pilot studies was to increase interest in and willingness to analyze and comprehend facial affect in a technology-aided conversational setting. The Emotion Bubbles interface was successful in helping the participants develop a fast intuitive understanding of what the Facial Analysis System interface was tracking and how different expressions might produce particular results. Through this new understanding, the participants were then able in conversations with their friends to try to cause particular emotion bubbles to increase in size. In particular, the experimenters witnessed multiple instances of individuals who were trying to elicit a particular expression adjusting their conversational flow to try to get that expression. According to the Groden Center staff, our pilots were also successful in helping the participants critically analyze facial expression in an attempt to find unspoken emotional cues.

In terms of interface design, it was illuminating to see how the individuals used the Emotion Bubbles in terms of their needs and preferences with regard to the graphical properties. As a result of their input, we added a number of features to the interface. The version used for our second round of pilots included features such as the ability to turn off each bubble individually (signaled by a series of bubbles in the bottom left; bubbles which were not currently being displayed were grayed out); the ability to move each bubble with the Samsung stylus; the ability to "freeze frame" to help analyze the Facial Analysis System interpretation of the last facial affect event; the capacity to turn the background black for easier viewing of the bubbles; and the ability to make the text larger and of different fonts to facilitate reading for individuals with visual and cognitive impairments that made the original text difficult to read. These considerations were difficult to anticipate prior to running the pilot study with these autistic individuals, and we were able to garner a significant amount of information about what accommodations should be considered in future interface design for people diagnosed with an ASD. The goals of simplicity and intuitiveness of design were never ignored; research in this field is supportive of the fact that autistic people do particularly well during analysis of information that is presented in such a way that enables core information to be easily discerned, leading to our emphasis on only showing colored bubbles for emotions which were visible at all times.

6. CONCLUSIONS

This paper describes new technology for just-in-time *in situ* learning of facial affect together with the results of a series of pilot studies conducted with adolescents diagnosed with ASD who used the technology in their peer-group setting and contributed to its development via their feedback. This is an important milestone in this area of research precisely because face-to-face, real-time understanding of emotional state is difficult for this population. By bringing "emotion" and facial affect from the realm of intangible concepts into the area of quantitatively measurable and analyzable information using the Facial Analysis System/Emotion Bubbles interface, we were able to allow these individuals to hold well-informed conversations about how their expressions were perceived by their friends and how they might change their facial expressions to demonstrate some particular idea.

Future interface design in this and related areas can benefit from the knowledge we obtained in these pilot studies. In particular, the principle that intangible information can be represented in a fun and intuitive way could be very helpful in multiple contexts for people with autism. This knowledge is particularly useful in light of the fact that interface design for people on the autism spectrum is an under-researched area that can benefit from increased attention.

The introduction of technology-assisted conversational and facial affect analysis may be extremely useful to autistic people in terms of helping them understand normative social interactions. Furthermore, science still doesn't have a deep understanding of social interaction (e.g., how often each facial expression tends to get reciprocated, what the statistics of natural eye contact are, etc.) The same tools developed to aid autistic people in this discovery process can be used to help advance the fundamental scientific understanding as well.

With regard to the hardware, our major goal is to streamline our system so that there are fewer and shorter wires and a simpler assembly process for camera/computer configurations. We may also add a tripod or an accelerometer that will allow the system to cope with "jitter" and process emotion accurately to counteract slight movement from the holder of the computer.

In our future work with the Groden Center, we will organize a monthlong pilot to measure the cumulative effects of working with the Emotion Bubbles interface. We will load the Facial Analysis System /Emotion Bubbles interface onto 3-5 Samsung ultramobile computers obtained for this purpose with and train the staff there to run our protocol, with modifications made as necessary for individuals of varying cognitive abilities and the availability of staff members. Throughout the course of the study, we will visit each of the 15 autistic for about 20 minutes a day, three days a week. During this time, we will work with the participants to help them think about their ability to communicate nonverbally and appreciate non-verbal communication with the assistance of the Facial Analysis System and the Emotion Bubbles. Our goal is to help this population become more confident in everyday circumstances with the idea of using and understanding emotion in conversational situations. We will assess the participants' general state with regard to emotion perception with the Texas Alexithymia Scale (TAS-20); we will also perform before-and-after assessments of the participants with the CAM Battery, the Classic Films Battery, and the Social Responsiveness Scale for the purpose of measuring the success of our approach in developing appropriate responses to communicated emotional state. Our goal is to make the mobile PCs a socially accepted part of interaction in that setting, so that the students will feel able to interact with each other and analyze their friends' expressions with the goal of enhancing effective communication and learning to better communicate through facial expressions.

7. ACKNOWLEDGMENTS
This material is based upon work supported by the National Science Foundation under Grant No. 0555411 and the Things That Think Consortium. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would like to especially thank the students and staff at the Groden Center for participating in our pilot. The hardware was generously provided by Samsung and Logitech, and the tracker used as a front-end for the face analysis software was provided by Google.

8. REFERENCES
[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR*. 2000.

[2] National Institute of Neurological Disorders and Stroke. *Autistic Disorder Fact Sheet*. December 12, 2007 [cited 2008 March 5]; Available from: http://www.ninds.nih.gov/disorders/autism/detail_autism.htm.

[3] National Institute of Neurological Disorders and Stroke. *Asperger's Syndrome Fact Sheet*. December 12, 2007 [cited 2008 March 5]; Available from: http://www.ninds.nih.gov/disorders/asperger/detail_asperger.htm.

[4] National Institute of Neurological Disorders and Stroke. *Autism Spectrum Disorders (Pervasive Developmental Disorders.)*. [cited 2008 March 5]; Available from: http://www.nimh.nih.gov/health/publications/autism/complete-publication.shtml.

[5] Bernard-Opitz, V., N. Sriram, and S. Nakhoda-Sapuan, *Enhancing social problem solving in children with autism and normal children through computer-assisted instruction.* Journal of Autism and Developmental Disorders, 2001. 31: p. 377-398.

[6] Swettenham, J., *Can children with autism be taught to understand false belief using computers?* Journal of Child Psychology and Psychiatry, 1996. 37: p. 157-165.

[7] Bolte, S., et al., *The development and evaluation of a computer-based program to test and to teach the recognition of facial affect.* International Journal of Circumpolar Health, 2002. 61(2): p. 61-68.

[8] Howlin, P., S. Baron-Cohen, and J. Hadwin, *Teaching Children with Autism to Mind-read: A Practical Guide for Teachers and Parents*. 1999: John Wiley and Sons.

[9] Rutherford, M.D., S. Baron-Cohen, and S. Wheelwright, *Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism.* Journal Autism and Developmental Disorders, 2002. 32: p. 189-194.

[10] Baron-Cohen, S., et al., *Mind Reading: The Interactive Guide to Emotions*. 2004, Jessica Kingsley Publishers: London.

[11] el Kaliouby, R. and P. Robinson, *Real-time Inference of Complex Mental States from Facial Expressions and Head Gestures*, in *Real-Time Vision for Human-Computer Interaction*. 2005, Springer-Verlag. p. 181-200.

[12] Zacks, J.M., B. Tversky, and G. Iyer, *Perceiving, Remembering, and Communicating Structure in Events.* Journal of Experimental Psychology, 2001. 130(1): p. 29-58.

[13] Ekman, P. and W.V. Friesen, *Facial Action Coding System*. 1977: Consulting Psychologist Press.

[14] Baron-Cohen, S., *The hyper-systemizing, assortative mating theory of autism.* Progress in Neuro-Psychopharmacology & Biological Psychiatry, 2006. 30: p. 865–872.

[15] Mottron, L., et.al., *Enhanced Perceptual Functioning in Autism*. Journal of Autism and Developmental Disorders, 2006. 30 (1): p. 27-43.