Emotion Recognition via Face Tracking with RealSense™ 3D Camera for Children with Autism

Tiffany Y. Tang
Media Lab
Dept. of Computer Science
Wenzhou-Kean University
yatang@kean.edu

Pinata Winoto
Media Lab
Dept. of Computer Science
Wenzhou-Kean University
pwinoto@kean.edu

Guanxing Chen
Media Lab
Dept. of Computer Science
Wenzhou-Kean University
ChenGuanxing316@wku.edu.cn

IDC '17, June 27-30, 2017, Stanford, CA, USA

ACM ISBN 978-1-4503-4921-5/17/06.
http://dx.doi.org/10.1145/3078072.3084321

Abstract
Although there is a growing recognition of the differences, not diminished abilities, of facial affective expressivity between Typically Developing (TD) and Autism Spectrum Disorder (ASD) individuals, which might lead to the varied recognizability of conveyed emotion by both TD and ASD individuals, little is explored on the ecological validity of these findings; that is, whether *spontaneous* affective facial expressions can better be produced and recognized by both populations. We aimed to address these issues in the present study, using children's cartoon clips to assess two aspects of spontaneous emotion production and recognition in a context closer to real-life children's cartoon movie watching (at home or a classroom). Based on the facial landmark data and a teacher/parent's manual emotion tags (happy), we performed a computational analysis to compare the happy emotion labels generated by the automated algorithm and the human TD rater. Two pilot studies of six ASD children revealed the potential as well as challenges of such an approach.

Author Keywords
Autism; emotion; children; affective expressivity; computational analysis; facial landmark; China

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

Introduction
One of the core characteristics of ASD is the presence of early and persistent impairments in social-communicative skills [1]. An overwhelming number of previous studies usually attribute such deficits to the population's diminished skills in emotion recognition until a few recent studies started to probe into possibility of the lack of such skills in the TD population. That is, awkward moments in social interactions might be characterized by facial emotion recognition difficulties on both sides of the social equation [4, 11, 22, 23] since social interactions are intrinsically bi-directional [11]; that said, instead of attributing the social-communicative skill deficits in ASD individuals to their impairments in emotion recognition skills, isn't it possible that TD people might also have difficulties or deficits at reading ASD expressions? Recent studies have indeed empirically demonstrated the difficulties of TD individuals in recognizing ASD expression [4, 7, 22]. Another related line of research has concluded that individuals with ASD demonstrate varied quality on *affective expressivity* across multiple modalities: specifically, for example, a greater variability in vocal prosody production [12, 16], the social inappropriateness of facial affective expressivity [3, 4, 7, 14, 22]. Some attributes the subtle differences of affective facial expressivity between the two populations to the reduced facial muscle movement [6, 15, 25], irregularities in limbic brain activities which are closely related to emotion [15], and being less motivated in engaging in social activities [4], in the ASD community.

These converging empirical evidences revealed the differences of facial affective expressivity between ASD and TD individuals might prevent the former from producing recognizable emotions to both populations [4, 7, 12, 22, 24] and thus further complicate their social interactions. A pressing and under-explored research path to pursue is how existing mature facial emotional algorithms could successfully label the emotion of expression posed by ASD individuals so as to provide social cues for their partners including both ASD and TD individuals. A few recent works started to explore down this avenue [14, 20] including our present one. Specifically, in present study, we attempted to examine the potential of computational emotion recognition of ASD children's facial expression collected during naturalistic tasks (cartoon-clip watching) by comparing the manual emotion tag from teachers/parents with the automatic one generated by our system.

Related Work
There is no lack of empirical studies examining the impairments of facial expressivity in individuals with ASD (among many, [13, 21]); however, few probe the degree of *recognizability* of these emotions by either ASD or TD raters [4]. These two indirect, yet related, lines of research will first be presented to motivate the present study.

The Affective Expressivity of Children with ASD
Previous research has argued that individuals with ASD demonstrate varied quality on affective expressivity across multiple modalities: specifically, for example, a greater variability in vocal prosody production [12], the social-inappropriateness of facial affective expressivity [3, 4, 7, 8, 9, 14, 22, 24]. For example, Faso et al [7]

Figure 1: Pilot testing environment setup in a private autism center.

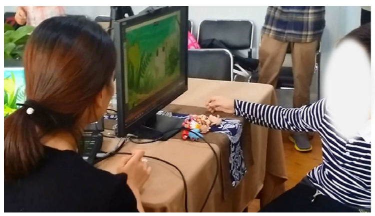


Figure 2: Pilot testing environment where a teacher sat side by side with the child to manually 'spot' his happy expression.

Figure 3: Another pilot testing in our lab where a girl was watching a cartoon movie.

examined how well facial expressions produced by individuals with ASD are perceived and thus subsequently recognizable by both TD and ASD individuals (both were referred to as the *potential social partners*). They have found that both groups of potential social partners did not find it difficult to identify the facial emotions posed by individuals with ASD; though posed expressions by ASD individuals are somewhat *different* and rated as '*awkward*'. Such awkwardness might make it difficult for TD individuals to interpret the emotion of their ASD peers, which in turn make it emotionally disconnect between the two populations. Similar observations on the awkwardness of affective expressivity by individuals with ASD were also reported in [8], and a computational comparative study in [14]. Meanwhile, prior research provided prevailing evidences that differences, not diminished abilities, are exhibited in affective facial expressivity ASD individuals [4, 7, 12, 24]. For instance, [12] reported that raters observed a higher frequency of laughter produced by ASD children than that of produced by TD children; adults with ASD can produce more recognizable anger emotion than happy one by naïve female observers [7]. Picard argued that individuals with ASD could exhibit incredible calmer facial expression (outward) than their true emotional state (inward, measurable through physiological data including Galvanic Skin Response or GSR) [17, 18], which largely supported these previous empirical and clinical findings.

Can TD Individuals Read the Emotion of Children with ASD?

Recent empirical studies began to probe the other side of the social interaction equation in observing whether the emotion of individuals with ASD can be recognized by TD individuals, including those have more intimate relationships with ASD individuals (parents, teachers, etc.) [4, 5, 22, 24] following the mounting evidences of unnatural and exaggerated affective facial expression posed by ASD individuals. In order to determine whether atypical emotion production is common or idiosyncratic, [4] investigated both TD and ASD judgers' ability to recognize emotional expressions produced by TD and ASD posers. Empirical studies are consistent with previous ones in highlighting the degree of recognizability decreased on expressions produced by ASD individuals regardless of the profiles of the raters which implicate idiosyncratic atypical emotional expression representation in ASD [4]. In a large sample of 84 ASD and TD children, empirical results indicated that TD adult judgers (senior psychology college students) were less able to recognize *sad* facial expression produced by children with ASD than that produced by TD children [24]. More recognizable *anger* emotion than *happy* one by adults with ASD was captured in [7]. ASD children were perceived, by parents, as showing more negative emotion than TD children.

Computational Facial Data Analysis for Emotion Labeling in Autism Research

Although emotion recognition has been an active research field, the majority of the earlier studies targeted and evaluated by typical users [26]. Among them, recognizing emotions from facial expression received the most attention; portable cameras make it possible to integrate the technique into (multimodal) everyday activities for emotion recognition [26]. Recently, the low-cost 3D sensor cameras such as Microsoft Kinect™ and the newer Intel RealSense™ receive much attention due to its advantages in

Facial Action Units (AUs)	Real-Sense	Facial Land-mark
Upper Eyelid	√	
Lower Eyelid	√	
Eyes Smaller		√
Corners of the Mouths	√	
Lip Medium	√	
Grin		√

Table 1: selected facial AUs for happy emotion in our study

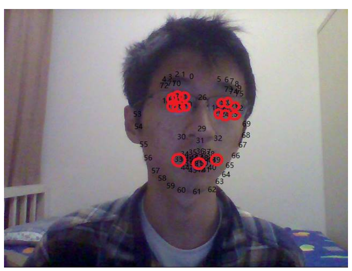


Figure 4 Selected facial landmarks obtained from RealSense

capturing fine-grained skeleton and facial landmark data which in turn provided motion data for more accurate analysis [20]. Such computational analysis has the potential to characterize emotion expressivity in terms of temporal feature of facial landmark data, head motion and hand gestures, especially with more affordable and portable motion and facial data capture technologies [23]. However, the inherently atypical facial muscle movement in individuals with autism poses significant challenges for such technique to be effective in the domain. Metallinou et al [14] quantitatively compared the affective facial poses of 37 individuals aged ranging from 9 to 14 with high functioning autism (HFASD, 21) and typically developing individuals (TD, 16). The task involved is to mimic the emotional facial expression from the common Mind Reading CD [2]. Motion capture technology and functional data analysis were deployed to allow the facial gesture data capture, analysis and visualization of temporal nature of facial landmark data. Experiential results revealed statistically significant differences between ASD and TD children in terms of the asynchrony of motion between different facial regions and more facial motion roughness [14]. [19] relied on human-intervened semi-automatic analysis and [20] revealed that for both TD and ASD children, happiness, sadness and anger were correctly labeled with high accuracy through Intel's RealSense™ lens. These recent works provided solid ground for our present study.

The System, Pilot Study, and Study Procedure

The evaluated system consists of a 19-inch LCD monitor, an Intel RealSense™ camera, and a computer with a typical camera (see Figure 1). The Intel RealSense™ SR 300 programmable camera was mounted on the LCD monitor where a two-minute cartoon was shown to capture the facial landmark data of the child; side by side with the RealSense™ is a typical high-definition camera to record their continuous facial expressions.

RealSense™ SDK could detect a total of 78 facial landmarks and calculate 16 facial action units (AUs) from a human face [20]. To detect the happy expression, six facial action units (AUs) are selected (see Table 1); Figure 4 illustrates an example of the selected facial landmarks used in our study. The linear distances between these landmarks were then computed every second to determine whether it is a happy or non-happy face [20]; that is, by comparing the distances with some discrimination thresholds.

Drawn from earlier results that caregivers and parents are likely to develop secure attachment with the ASD individuals [5] and thus can easily spot the happy and anxious moments in an ASD child [10], we asked either a home-run teacher or parent to sit in front of another computer to 'press' the 'enter' key when s/he spotted a happy moment in the child (see Figure 2). We assume the observation made by teachers/parents is more accurate than that from our evaluated system in judging their children's happy emotion, because they could observe not only children's facial expression, but also their voice, gesture, and other contextual information (multi-modalities).

Finally, we manually compared the labels ('happy') obtained from the evaluated system with the labels provided by teachers/parents. The maximum temporal

Data Collection
Before the testing, a pre-test interview with either teachers or parents was conducted on some basic statistics of the child such as the age of children, the schooling time, their self-reported ability on recognizing the happy emotion of the child, relative frequency of happiness spotted, etc. During the testing, we asked children to watch a two-minute cartoon in front of the display monitor with a RealSense™ camera attached on the top of it. Meanwhile, a teacher or parent pressed 'Enter' key when the child's happy emotion was spotted and automatically recorded (see Figure 2).

offset of the co-occurrence is set to three seconds; that is, an agreement between the system and human is reached if the system could detect a happy face within three seconds before/after the teachers/parents press the enter key. Any consecutive 'happy' labels from the system would be considered as a single happy moment.

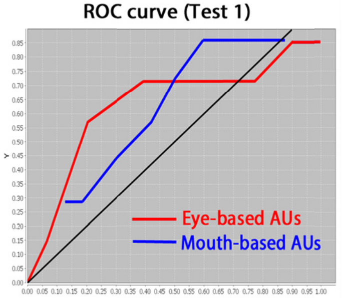
Pilot Study Results and Discussion

The experiments were carried out twice (i.e. in a private autism center and our lab, see Figure 2 and 3 respectively). The pre-test interview revealed that both teachers and parents can tell when the child is happy and the frequency of happiness is relatively satisfying. A total of six ASD children (Mean age=4.5, SD=1.2) taking part in the testing and the system had been tested ten times. Among them, only six testing moments from four ASD children are valid. Invalid data are caused by either teachers/parents' failure in operating the equipment or due to children's temporary disruptive behaviors. Using the collected valid data, a discrimination threshold δ is adjusted in the computation of AUs in our analysis. By decreasing the threshold value, our system becomes more sensitive in detecting a happy face; hence, increasing both its true positive and false positive rates, and vice versa. Here, the true positive rate refers to the probability of detecting a happy emotion correctly by the system, and the false positive rate refers to the probability of misjudging a non-happy emotion as a happy one. Those said, we can draw the Receiver Operating Characteristic (ROC) curve after adjusting δ. These ROC curves represent the accuracy of the evaluated system in recognizing children's facial expression (happiness) relative to their teachers/parents' observation. The horizontal axis of an ROC curve represents the false positive rate and the vertical axis represents the true positive rate (see Figure 5). The origin (0, 0) represents the situation when the system is extremely insensitive in recognizing happy face; hence, it would never classify a face as a happy one regardless the teachers/parents' observation. In contrast, the point (1, 1) represents the situation when the system is oversensitive or would always classify a face as a happy one. Finally, the point (0, 1) represents the ideal system which could always correctly classify happy/non-happy emotion exactly as what the teachers/parents have observed. Figure 5 shows two sets of ROC curves for eye-based AUs and mouth-based AUs. A diagonal black line represents a random classifier; the further a point is above it the better our classifier is. It appears in Figure 5 that both eye-based and mouth-based AUs could be used to predict children's happiness, where eye-based AUs are more stable as a predictor. This is true, because when we fix the false positive rate at 0.3, the average true positive rates from all six samples are 0.69 and 0.60 for eye-based and mouth-based AUs, respectively.

Concluding Remarks and Challenging Issues

In present study, we attempted to examine the potential of computational emotion recognition of ASD children's facial expression collected during naturalistic tasks (cartoon-clip watching) by comparing the manual emotion tag from teachers/parents with the automatic one generated by our system. Experimental results favor the use of facial landmark data around the eye area. More experiments need to be conducted to further examine the validity of it.

In the long run, such tools could provide assistance to help TD individuals to read the emotions of ASD. There existing two key challenges along this research avenue,



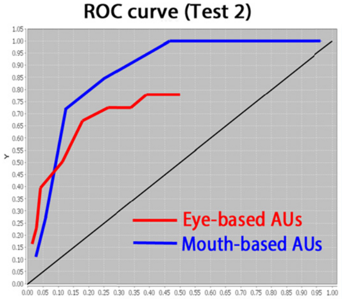


Figure 5. ROC curves of two children using eye-based AUs and mouth-based AUs respectively.

all of which are rooted in the *idiosyncratic*, rather than common and systematic, representation of emotion expression in people with ASD [4]. Specifically, how to validate the two emotion labels generated by the computer algorithm and manually tagged by a human user; and how to *groupize* and *adapt* such an automatic emotion recognition tool to at least a *subtype* of ASD individuals remain key open questions as well as our future works.

Acknowledgements

We would like to thank all parents, children and teachers involved in the study; thanks also go to Xi Yang, Xiaoyang Qiu, and Hang Xie for their assistances in the pilot testing. The financial support of this research through Student Partnering with Faulty (SpF) Program (WKU201617004) is greatly acknowledged.

References

1. American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington DC: American Psychiatric Association.
2. Simon Baron-Cohen. 2003. *Mind Reading: The Interactive Guide to Emotions*. Jessica Kingsley Publisher, London, UK.
3. Sander Begeer, Hans M. Koot, Carolien Rieffe, Mark Meerum Terwogt, and Hedy Stegge. 2008. Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review* 28,3: 342–369.
4. Rebecca Brewer, Federica Biotti, Carloline Catmur, Clare Press, Francesca Happé, Richard Cook, and Geoffrey Bird. 2016. Can Neurotypical Individuals Read Autistic Facial Expressions? Atypical Production of Emotional Facial Expressions in Autism Spectrum Disorders. *Autism Research* 9 2: 262–271.
5. Lisa Capps, Connie Kasari, Nurit Yirmiya, and Sigman, M. 1993. Parental perception of emotional expressiveness in children with autism. *J. Consulting and Clinical Psychology* 61, 3: 475–484.
6. Patrycja Czapinski, Susan E Bryson. 2003. Reduced facial muscle movements in autism: Evidence for dysfunction in the neuromuscular pathway? *Brain and Cognition* 51, 2: 177–179.
7. Daniel J. Faso, Noah J. Sasson, Amy E. Pinkham. 2015. Evaluating Posed and Evoked Facial Expressions of Emotion from Adults with Autism Spectrum Disorder. *J. Autism and Developmental Disorders 45, 1:* 75–89.
8. Ruth B. Grossman, Lisa R. Edelson, and Helen Tager-Flusberg. 2013. Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism. *Journal of Speech, Language and Hearing Research* 56, 3: 1035.
9. Ruth B Grossman. 2015. Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. *Autism* 19, 5: 580-587.
10. Amanda C. Gulsrud, Laudan B. Jahromi, and Connie Kasari. 2010. The Co-regulation of emotions between mothers and their children with autism. *J. Autism and Developmental Disorders* 40, 2: 227–237.
11. Amy G. Halberstadt, Susanne A. Denham, and Julie C. Dunsmore. 2001. Affective Social Competence. *Social Development* 10, 1:79–119.
12. William J Hudenko, ad M. A. Magenheimer. 2011. Listeners prefer the laughs of children with autism to those of typically developing children. *Autism 16, 6:* 641–665.
13. Connie Kasari, Marian Sigman, Peter Mundy, and Nurit Yirmiya. 1990. Affective sharing in the context of joint attention interactions of normal,

autistic, and mentally retarded children. *J. Autism and Developmental Disorders*, 20(1), 87–100.

14. Angeliki Metallinou, Ruth B. Grossman, and Shrikanth Narayanan. 2013. Quantifying atypicality in affective facial expressions of children with autism spectrum disorders. In *Proc. of 2013 IEEE International Conference on Multimedia and Expo* (ICME). DOI: 10.1109/ICME.2013.6607640
15. Rachel L. Moseley, Yury Shtyrov, Bettina Mohr, Michael V. Lombardo, Sim on Baron-Cohen, Friedemann Pulvermüller. 2015. Lost for emotion words: What motor and limbic brain activity reveals about autism and semantic theory. *NeuroImage* 104: 413–422.
16. Aparna Nadig, Holly Shaw. 2012. Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means for listeners. *Journal of Autism and Developmental Disorders* 42, 499–511.
17. Rosalind Picard. 2009. Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B*, 364:3575-3584.
18. Rosalind Picard. 2010. Emotion research by the people, for the people. *Emotion Review 2, 3:* 250-254.
19. James M. Rehg, Agata Rozga, Gregory D. Abowd and Matthew S. Goodwin. 2014. Behavioral Imaging and Autism, *IEEE Pervasive Computing*, 13, 2:84-87.
20. Vinícius Silva, Filomena Soares, João Sena Esteves, Joana Figueiredo, Cristina Santos and Ana Paula Pereira. 2017. Happiness and Sadness Recognition System—Preliminary Results with an Intel RealSense 3D Sensor. In *Proc. of CONTROLO 2016*, LNCS 402, Springer.
21. Margaret E. Snow, Margaret E. Hertzig and Theodore Shapiro. 1987. Expression of Emotion in Young Autistic Children. *J. the American Academy of Child and Adolescent Psychiatry* 26, 6: 836–838.
22. Steven D. Stagg, Rachel Slavny, Charlotte Hand, Alice Cardoso, and Pamela Smith. 2014. Does facial expressivity count? How typically developing children respond initially to children with autism. *Autism* 18, 6: 704–711.
23. Tiffany Y. Tang. 2016. Helping Neuro-typical individuals to read the emotion of children with autism spectrum disorder: an internet-of-things approach. In *Proc. of the 15th ACM Interaction Design and Children Conference (ACM IDC2016),* 666-671.
24. Martin A. Volker, Christopher Lopata, Donna A. Smith, Marcus L. Thomeer. 2009. Facial encoding of children with high-function autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities* 24, 4: 195–204.
25. Amy K. Weimer, Amy M. Schatz, Alan Lincoln, Angela O Ballantyne, and Doris A. Trauner. 2001. “Motor” impairment in asperger syndrome: Evidence for a deficit in proprioception. *Journal of Developmental & Behavioral Pediatrics* 22,2: 92–101.
26. Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31, 1:39–58.