Mobile Crowdsourcing of Emotive Video for Autism Research

Haik Kalantarian
Stanford University
Stanford, CA 94305
haik@stanford.edu

Peter Washington
Stanford University
Stanford, CA 94305
peter100@stanford.edu

Jessey Schwartz
Stanford University
Stanford, CA 94305
jesseys@stanford.edu

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

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

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

Abstract

Autism Spectrum Disorder (ASD) is a condition affecting 70 million children worldwide. Due to delays in diagnosis and imbalances in coverage, it is necessary to develop new methods of care delivery for use outside of clinical settings. In this paper, we present a mobile charades-style game, *Guess What?*, used for crowdsourcing the acquisition of structured video from children with ASD for behavioral disease research. By analyzing facial affect in response to various prompts, we demonstrate that engagement and facial affect can be quantified and measured using real-time image processing algorithms: an important first-step for future therapies, diagnostics, and outcome measures based on home video. Our study of thirteen subjects reveals that game sessions displaying *faces* to the player produced the most emotive facial expressions in the player by a considerable margin. Videos from the younger neurotypical group contained 73.9% more frames with emotion in this category, compared to the older group with ASD. Moreover, results from all categories indicated 20.7% less facial engagement among older children with ASD compared to younger children with no diagnosis. We will attempt to replicate these findings in a larger cohort in future work.

Author Keywords
Autism, Mobile, Crowdsourcing, Computer Vision, Emotion

ACM Classification Keywords
Applied Computing [Life and Medical Sciences]: Health Informatics.

Introduction
Autism spectrum disorder (ASD) is a developmental disability that affects 70 million children worldwide, including 750 thousand American children under the age of ten [25, 7]. The prevalence of ASD has increased over the years, with the rate of incidence climbing from 1 in 125 to 1 in 68 US children within the last 5 years [25, 7]. Most children with ASD struggle to make eye contact, recognize facial expressions, and engage in social interactions with their family and peers [19, 21, 12]. Standard methods of ASD diagnosis include the Autism Diagnostic Observation Schedule (ADOS) [23] and Autism Diagnosis Instrument Revised (ADI-R)[24]. These instruments are delivered in clinics and require hours of administration by specialists. The long length of standard exams as well as the need for administration in a clinical facility contribute to delays in diagnosis and an imbalance in coverage of the population needing attention. There is an urgent need to develop new methods of care delivery that can appropriately empower children and caregivers.

We present *Guess What?*: a mobile game based on the mechanics of the popular charades game, *Heads Up*. In *Guess What?*, children act out various images shown on the phone while being recorded by their parents, who attempt to guess the prompt. The parent labels the video by tilting the phone after they have correctly identified what their child is expressing, at which time another

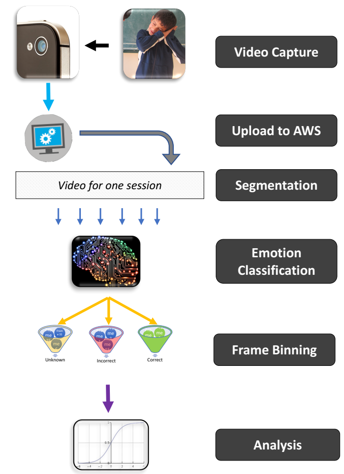

Figure 1: Frames from video captured during a game session are binned by emotion. Various prompts are explored for their efficacy in evoking emotive facial expressions.

prompt is presented until time is up. Using this method of crowdsourced at-home video acquisition, we are developing a computer vision database of children with ASD as well as neurotypical children as they express themselves in response to various prompts: emojis, facial expressions, and more abstract prompts such as sports games, occupations, and animals. Our prior work [30][8][29] has demonstrated the efficacy of mobile video phenotyping approaches for children with ASD. Similarly, the game mechanics of *Guess What?* are well-suited for ASD research because they facilitate video capture in a semi-structured environment in which the child's face is almost always centered correctly in the frame. The system architecture is shown in Figure 1; our system collects videos as children react to various prompts shown on the phone while being recorded using the device's front camera. Subsequently, facial-expression recognition algorithms [18] are employed to analyze video frames individually, recognizing emotion and monitoring facial engagement in real-time during the game session.

Related Work

A large body of work in recent years has explored the use of technology to enhance therapy for children with Autism Spectrum Disorder (ASD). For example, one study revealed that watching a television series designed to enhance emotion comprehension on a daily basis can improve the emotion comprehension of young children with ASD [17]. An example of a novel therapy is the ECHOES project, which presents a technology-enhanced learning (TEL) environment that facilitates acquisition and exploration of social skills by typically developing children and children with ASD [5]. Multitouch interfaces for behavioral therapy have been found to increase collaboration skills in children with ASD [4, 15, 16]. Frameworks for designing games to teach emotion to children with ASD have also been shown to be successful [26, 29, 30]. Others have explored the use of wearable systems and affective computing as companion tools for social-emotional learning and the use of the recorded videos for defining a process to collect, segment, label, and use video clips from everyday conversations [27, 28].

System Description

After account creation, parents are directed to the deck-selection screen. The possibilities include: emoji, animals, faces, sports, occupations, and a random category which selects among all of these. Naturally, the choice of prompts dramatically affects the quality of the incoming video data; showing the child photos of real people is asking them to imitate the expression within the

Figure 2: During a gameplay session, the screen is oriented towards the child. The child will see a prompt and will act it out using gestures and expressions while being recorded.

photo, while showing a more abstract image such as that of an astronaut challenges their creativity and interpretation skills. During the 90-second game session, the child will attempt to act out the prompt shown on the device while the parent guesses. After the parent guesses correctly, they will tilt the phone forward to indicate that the child has earned a point; the image will then change and the process will repeat until time is up. If the user elects to share the footage, it is added to a local buffer and uploaded to a secure and encrypted IRB-approved Amazon Web Services S3 bucket in a background process when the user connects to WiFi.

Video frames are processed using a face tracker algorithm which attempts to locate the face (if any) within the frame. When the child looks away from the camera during a game session, it is likely that the face tracker algorithm will fail. Therefore, the percentage of frames within a video in which the face tracker fails can be used as a general heuristic for a lack of facial engagement by the child during the game. For those frames in which the face tracker is successful, a Histogram of Oriented Gradients technique is used in conjunction with a linear SVM classifier to classify between eight classic emotions. These

Table 1: Summary of Two Groups of Participants

Group	Subjects	Avg. Age	Std. Dev	ASD
Group A	8	8.5 years	1.85 years	Yes
Group B	5	4.4 years	0.54 years	No

eight emotions include the seven universal emotions identified by Eckman et al. in [14], as well as a neutral category. The number of face tracker failures, percentage of neutral frames, and the percentage of emotive frames in which the face tracker was successful, are all recorded and aggregated on a per-subject and per-deck basis. Our implementation of the emotion classifier used in this work is based on our prior work by Haber et al. in [18].

Results

A total of thirteen subjects participated in lab sessions in which they played *Guess What?* on a Google Pixel phone running Android 7.1. To ensure consistency, the same individual, a member of our research staff, administered each game session. Participants played up to five games with the following decks: emoji, faces, animals, sports, and jobs; in several other cases, the impairment or age of the child made playing the more complex decks (occupation, sports) infeasible.

Participant demographic information is provided in Table 1. The average age of participants was 6.9 years $\pm$ 2.53. From these subjects, a total of eight were diagnosed with ASD (*Group A*), while the remaining five were neurotypical (*Group B*). The age difference between the ASD and neurotypical group was substantial: the latter was 8.5 years $\pm$ 1.85, while the former was 4.4 years $\pm$ .54.

Conclusion

This study has yielded a number of interesting conclusions. First, we learned that the most emotive responses are associated with the *faces* category by a considerable margin, followed by the emoji category. Jobs and sports expectedly produced very little recognizable emotion. The easiest category was animals, as children almost always expressed the prompt via vocalizations. By contrast, the jobs and sports categories are the most difficult. The easiest specific prompts are basic animals: pig, frog, lion, sheep, as well as several faces such as disgust and surprise. The most difficult prompts are sports outside of basketball, baseball, and football, as well as various occupations (dentist, astronaut).

We also discovered that videos from the faces category were much more efficacious in producing emotional facial expressions in children, with far less emotion derived from emoji-based videos. The most substantial differences between the two tested populations were also observed in the faces category. While performance between groups was similar across most categories, the neurotypical children scored more points in the faces category despite being younger. Video the younger neurotypical group contained 73.9% more frames with emotion in this category. Moreover, data derived from videos of children in Group A indicated 20.7% lower amount of facial engagement across all categories.

Acknowledgements

We are grateful to the study participants and for funding support to Dennis P. Wall from NIH (1R01EB025025-01 & 1R21HD091500-01), The Hartwell Foundation, and a program grant from Stanford's Precision Health and Integrated Diagnostics Center. We also acknowledge generous support from Bobby Dekesyer and Peter Sullivan.

REFERENCES

1. Autism society: What is autism? `http://www.autism-society.org/what-is/`. Accessed: 2017-010-30.
2. Guess what? - android apps on google play.
3. Heads up - android apps on google play.
4. Battocchi, A., Pianesi, F., Tomasini, D., Zancanaro, M., Esposito, G., Venuti, P., Ben Sasson, A., Gal, E., and Weiss, P. L. Collaborative puzzle game: a tabletop interactive game for fostering collaboration in children with autism spectrum disorders (asd). In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ACM (2009), 197–204.
5. Bernardini, S., Porayska-Pomsta, K., and Smith, T. J. Echoes: An intelligent serious game for fostering social communication in children with autism. *Information Sciences 264* (2014), 41–60.
6. Bernier, R., Mao, A., and Yen, J. Psychopathology, families, and culture: autism. *Child and adolescent psychiatric clinics of North America 19*, 4 (2010), 855–867.
7. Christensen, D. L., Bilder, D. A., Zahorodny, W., Pettygrove, S., Durkin, M. S., Fitzgerald, R. T., Rice, C., Kurzius-Spencer, M., Baio, J., and Yeargin-Allsopp, M. Prevalence and characteristics of autism spectrum disorder among 4-year-old children in the autism and developmental disabilities monitoring network. *Journal of Developmental & Behavioral Pediatrics 37*, 1 (2016), 1–8.
8. Daniels, J., Schwartz, J., Haber, N., Voss, C., Kline, A., Fazel, A., Washington, P., De, T., Feinstein, C., Winograd, T., et al. 5.13 design and efficacy of a wearable device for social affective learning in children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry 56*, 10 (2017), S257.
9. Dawson, G. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and psychopathology 20*, 3 (2008), 775–803.
10. Dawson, G., and Bernier, R. A quarter century of progress on the early detection and treatment of autism spectrum disorder. *Development and psychopathology 25*, 4pt2 (2013), 1455–1472.
11. Dawson, G., Jones, E. J., Merkle, K., Venema, K., Lowy, R., Faja, S., Kamara, D., Murias, M., Greenson, J., Winter, J., et al. Early behavioral intervention is associated with normalized brain activity in young children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry 51*, 11 (2012), 1150–1159.
12. Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., and Varley, J. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics 125*, 1 (2010), e17–e23.
13. Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., and Varley, J. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. *Pediatrics 125*, 1 (2010), e17–e23.
14. Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology 53*, 4 (1987), 712.

15. Gal, E., Bauminger, N., Goren-Bar, D., Pianesi, F., Stock, O., Zancanaro, M., and Weiss, P. L. T. Enhancing social communication of children with high-functioning autism through a co-located interface. *Ai & Society 24*, 1 (2009), 75.

16. Giusti, L., Zancanaro, M., Gal, E., and Weiss, P. L. T. Dimensions of collaboration on a tabletop interface for children with autism spectrum disorder. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011), 3295–3304.

17. Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V., and Baron-Cohen, S. Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. *Journal of autism and developmental disorders 40*, 3 (2010), 269–279.

18. Haber, N., Voss, C., Fazel, A., Winograd, T., and Wall, D. P. A practical approach to real-time neutral feature subtraction for facial expression recognition. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (March 2016), 1–9.

19. Howlin, P., Goode, S., Hutton, J., and Rutter, M. Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry 45*, 2 (2004), 212–229.

20. International, W. B. Heads up. `https://play.google.com/store/apps/details?id=com.wb.headsup`. Accessed: 2017-010-30.

21. Landa, R. J., Holman, K. C., and Garrett-Mayer, E. Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of general psychiatry 64*, 7 (2007), 853–864.

22. LoBue, V., and Thrasher, C. The child affective facial expression (cafe) set: Validity and reliability from untrained adults. *Frontiers in psychology 5* (2015), 1532.

23. Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., Bishop, S., et al. *Autism diagnostic observation schedule: ADOS-2*. Western Psychological Services Los Angeles, CA, 2012.

24. Lord, C., Rutter, M., and Le Couteur, A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders 24*, 5 (1994), 659–685.

25. Newschaffer, C. J., Croen, L. A., Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., Mandell, D. S., Miller, L. A., Pinto-Martin, J., Reaven, J., et al. The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health 28* (2007), 235–258.

26. Park, J. H., Abirached, B., and Zhang, Y. A framework for designing assistive technologies for teaching children with asds emotions. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, ACM (2012), 2423–2428.

27. Porayska-Pomsta, K., Frauenberger, C., Pain, H., Rajendran, G., Smith, T., Menzies, R., Foster, M. E., Alcorn, A., Wass, S., Bernadini, S., et al. Developing technology for autism: an interdisciplinary approach. *Personal and Ubiquitous Computing 16*, 2 (2012), 117–127.

28. Teeters, A. C. *Use of a wearable camera system in conversation: Toward a companion tool for social-emotional learning in autism*. PhD thesis, Massachusetts Institute of Technology, 2007.

29. Voss, C., Washington, P., Haber, N., Kline, A., Daniels, J., Fazel, A., De, T., McCarthy, B., Feinstein, C., Winograd, T., et al. Superpower glass: delivering unobtrusive real-time social cues in wearable systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, ACM (2016), 1218–1226.

30. Washington, P., Voss, C., Kline, A., Haber, N., Daniels, J., Fazel, A., De, T., Feinstein, C., Winograd, T., and Wall, D. Superpowerglass: A wearable aid for the at-home therapy of children with autism. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1*, 3 (2017), 112.

31. Wiggins, L. D., Baio, J., and Rice, C. Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Journal of Developmental & Behavioral Pediatrics 27*, 2 (2006), S79–S87.