A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study

Jordan Hashemi[*1], Kathleen Campbell[*2], Kimberly L.H. Carpenter[*2,3], Adrianne Harris[*4], Qiang Qiu[1], Mariano Tepper[1], Steven Espinosa[1], Jana Schaich Borg[2,3], Samuel Marsan[2,3], Robert Calderbank[1], Jeffery P. Baker[2,5], Helen L. Egger[2,3], Geraldine Dawson[2,3], and Guillermo Sapiro[1]

[1] Department of Electrical and Computer Engineering, Duke University, USA.
[2] School of Medicine, Duke University, USA.
[3] Department of Psychiatry and Behavioral Sciences, Duke University, USA.
[4] Department of Clinical Psychology, Duke University, USA.
[5] Department of Pediatrics, Duke University, USA.
[*] indicates co-first authors

ABSTRACT

In spite of recent advances in the genetics and neuroscience of early childhood mental health, behavioral observation is still the gold standard in screening, diagnosis, and outcome assessment. Unfortunately, clinical observation is often subjective, needs significant rater training, does not capture data from participants in their natural environment, and is not scalable for use in large populations or for longitudinal monitoring. To address these challenges, we developed and tested a self-contained app designed to measure toddlers' social communication behaviors in a primary care, school, or home setting. Twenty 16-30 month old children with and without autism participated in this study. Toddlers watched the developmentally-appropriate visual stimuli on an iPad in a pediatric clinic and in our lab while the iPad camera simultaneously recorded video of the child's behaviors. Automated computer vision algorithms coded emotions and social referencing to quantify autism risk behaviors. We validated our automatic computer coding by comparing the computer-generated analysis of facial expression and social referencing to human coding of these behaviors. We report our method and propose the development and testing of measures of young children's behaviors as the first step toward development of a novel, fully integrated, low-cost, scalable screening tool for autism and other neurodevelopmental disorders of early childhood.

Keywords

Autism, automatic behavioral coding, facial affect coding system, integrated app, scalability, natural environments.

1. INTRODUCTION

Early identification of children with impairing neurodevelopmental and psychiatric disorders, such as autism spectrum disorders (ASD) and anxiety disorders, can have a large impact on long-term development and outcomes [3, 4]. However, the social communication and affective behaviors associated with these disorders are difficult to measure and monitor in primary care pediatric settings, where children are commonly seen, and even more difficult to measure in natural environments such as homes or schools. Currently, identification of children with neurodevelopmental disorders in early childhood requires low specificity screening with questionnaires followed by time-consuming observational assessments by highly-trained clinicians. Even in the field of autism, where progress has been made in developing new technologies for measuring risk for the disorder, there is not yet a low-cost and scalable way to directly observe children and characterize their development automatically and on a ubiquitous mobile device [9, 19]. There is a need for novel, scalable tools to measure social communication and affective behaviors and to use such measures to identify children who are at risk for neurodevelopmental disorders in a manner that can be disseminated to primary care clinics, schools, or homes. The use of mobile devices is one way to address these challenges, since their popularity opens the door to low-cost (software only), training-free, and scalable solutions.

Towards this goal, we developed a mobile application to elicit and quantify social referencing and affective behaviors. These behaviors, which include reciprocal social interactions such as social smiling, social referencing, pointing, and directing facial expression to others, represent some of the earliest signs of autism [2, 18]. We investigated the concept of delivering visual stimuli that could elicit and analyze social referencing and affective behaviors of young children. We tested tablet-based delivery of the app with integrated stimuli presentation and behavior recordings in a 5-minute session during well-child medical visits and analyzed resulting data. We present our methodology and results of feasibility analysis for automated coding of these behaviors from children with and without autism.

MOBIHEALTH 2015, October 14-16, London, Great Britain

DOI 10.4108/eai.14-10-2015.2261939

2. METHODOLOGY

2.1 Experimental setup

The study was carried out in a pediatric care clinic and our laboratory, with the approval of the Duke Institutional Review Board. A researcher approached parents of children presenting for an 18 or 24-month well-child check at the end of their visit and after the child had been screened for autism as per usual care. Children with known hearing or vision impairments and parents who could not complete consenting in English were excluded. We used a simple set up of an iPad and stand. The parents held the child on his or her lap and the iPad was set about 1 meter away. To minimize distraction, other adults and children were asked to stand behind the parent. This also forced the child to turn to direct social communication to others for unambiguous coding of behaviors. Parents were told they could interact with their child for the first 45 seconds of the experiment while a mirror was presented on the screen, but then they were asked to remain quiet and not direct their child's behavior or attention once video stimuli began (unless the child became distressed or tried to get up). The examiner called the child's name at three prompted points during the stimuli and silently waved if the child turned and made eye contact. The frontal camera in the iPad recorded video throughout the stimuli presentations at 1280x720 resolution and 30 frames per second. If children screened positive on standard autism screening with the Modified Checklist for Autism in Toddlers - Revised (M-CHAT-R), or a parent or clinician expressed concerns about autism during the medical visit, children received gold standard diagnostic testing with the ADOS-T with a child psychologist for final diagnosis and group assignment [13, 20].

Data has been collected on 47 children at the time of this analysis. Two children refused to finish watching the videos and one child did not have sufficient landmarks for analysis due to hair covering his face. We selected 20 subjects for the feasibility study based on age distribution to represent the range of ages and both diagnostic groups. Data from the first two video stimuli (bubbles and a social scene featuring a mechanical bunny, Figure 1) were selected for this feasibility analysis because preliminary analysis on pilot subjects showed that smiling and social referencing (looking at parent or examiner) was elicited most reliably during this part of the experiment. Total time analyzed was 97 seconds. We chose frequency of positive affect and frequency of social referencing as the main outcome variables, as these will be our clinically important outcome variables for future group analysis comparing children with and without autism. We removed the ratings of surprise in both human and computer coding due to frequent vocalizations in young children which cause the mouth to open briefly and the tendency of children in this age group to stare with mouth slightly open. For feasibility analysis, 6 children with autism and 14 age-matched controls were processed through automatic and human coding of videos. To determine feasibility of elicitation of social communication behaviors and automated coding, we compared results of automated and human coding in the combined dataset of autistic and control subjects.

2.2 Computer coding

Computer vision approaches for facial analytics rely on extracting features around specific regions on a face, such as the mouth, eyes, and nose [21, 22, 23, 24, 28]. We classify facial expressions and track head pose from 49 facial landmarks automatically extracted via the IntraFace software [26]. For the task of facial expression classification we employ a modified version of the method previously described in Hashemi et al [8], which is a robust method shown to handle expression classification across varying poses. This is critical for analyzing young children's natural behavior, since more standard approaches that constrain the user would not be appropriate for use with young children. We learn a cross-modality and pose-invariant dictionary using the BU-3D Facial Expression dataset [27]; and train the facial expression classifier based on the learned dictionary and the standard Cohn-Kanade dataset [11], where 3 classes are considered: Neutral, Positive (Happy), and Negative (Anger, Disgust, and Sad). For the task of head pose, we incorporate the pose output from the IntraFace software [26, 10]. Given a video stream comprised of consecutive images (frames), we first classify an emotion label and estimate head pose for each frame independently, and then we smooth the emotion labels by applying a max-voting filter every half of a second (15 frames). If in any given frame there is not a visible face, or the face exhibits a drastic pose relative to the camera ($> 45^\circ$ or $< -45^\circ$ yaw pose), the algorithm returns a 'Not Visible' tag on that frame. Since the parent and examiner are behind the child, an instance of social referencing is defined by a period in time when the yaw pose changes from $\leq 45^\circ$ to $> 45^\circ$, and then comes back a period of time later while exhibiting a pose $< 45^\circ$ and $> 35^\circ$. Similarly, social referencing also considers the case when the yaw pose changes from $\geq -45^\circ$ to $< -45^\circ$, and back to $> -45^\circ$ and $< -35^\circ$. These head pose values were chosen since our head pose algorithm requires both eyes to be visible on the face and to reflect what human coders will consider clear social referencing.

Figure 1: Example images of the experimental setup. From left to right, the first two images show screenshots from the two video-stimuli. The right-most image is an example of the recording acquired from the front camera on the iPad.

2.3 Human coding

To classify facial expression, we used the principles of anatomic units from the baby Facial Action Coding system (BabyFACs) of affect coding and rated only hedonic tone (positive, negative, neutral) to maximize reliability and generalizability [1, 15]. We chose the more general system of coding hedonic tone because previous studies have shown that observers can classify hedonic tone on young children with high accuracy [17]. Briefly, raters coded a positive expression when the action of the zygomaticus major pulled the lip corners up, negative tone when the action of the corrugator supercilli caused brow lowering, surprise when the mouth was wide open, and neutral affect when none of these muscle movements were present [12]. Raters coded 'Not Visible' when child's face was covered or out of the field of view and when more than half of the face was not visible due to head turning away from the camera. Social referencing was coded

when a child turned to look at the parent or examiner who was behind them. Coding was performed in Nodlus Observer XT software version 11.0 [16]. Raters first trained on a reliability dataset (separate from analyzed subjects) until they reached agreement greater than 75%. A single rater coded the dataset of subjects for these analyses; a second rater coded 20% of the dataset to verify on-going inter-rater reliability. Inter-rater agreement for total time when raters gave the same code to a behavior was 84% (76%-95% range). Raters were not blind to diagnostic group; but were blind to stimuli and videos were muted during coding to prevent the influence of vocalizations on the coding of hedonic tone.

3. RESULTS

3.1 Agreement for affect

We compared frame-by-frame behavior coding between the human coder and the automatic classifier to determine how much time the computer gave the same code as a human rater (example in Figure 2). This time resolution is significantly more accurate than what is done in clinical screening. Time in agreement and % agreement were calculated for each child, showing a range of 30-96% agreement with a mean of 75% (Table 1). The outlying value of 30% (subject 8) was due to disagreement between human coding of negative and computer coding of neutral. Some of the disagreement was due to edge effects, where the human and computer agreed on a prolonged facial expression, but did not agree on small changes in expression and on the exact frame-by-frame onset and offset of an expression. Additionally, we tested inter-rater reliability on frequency of positive affect as a potential outcome measure by calculating the intraclass correlation coefficient (ICC) using the package 'irr' in R [5, 25]. We quantified frequency by extracting the distinct instances of positive affect lasting greater than 0.5 seconds to limit measurement of small movements in the child's face, which represent noise due to the high sampling rate rather than true expressions [14]. We used a two-way, consistency, average measure ICC [7]. ICC for frequency of positive affect was 0.69 (95% CI 0.21-0.88), reaching good agreement, but with a wide confidence interval, suggesting variance between raters for some subjects. We expect smoothing of data to a higher sampling rate and disregarding insignificantly brief codes will generate better agreement in future analysis, and be more representative of the clear agreement like that observed in the visualizations in Figure 2.

Figure 2: Time-based coding by automated computer (top) and human (bottom) methods demonstrating high agreement (subject 9; stimulus 1) with 27 out of 30 seconds agreement. Green is neutral expression, red is positive expression, black is not visible, and light green is social referencing (looking at adult).

Table 1: Human coding compared to automated coding showing good overall agreement (75% across 20 subjects) and no apparent difference in mean agreement between the autism (A) and control (C) groups.

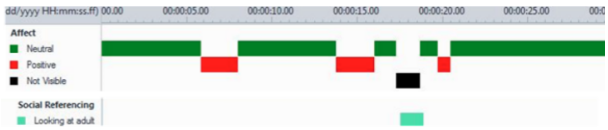
Subject (diag.)	Age (months)	Agree. (seconds)	Disagree. (seconds)	% Agree.
1 (A)	17	80	17	83%
2 (A)	20	77	20	79%
3 (A)	25	63	34	65%
4 (A)	30	84	13	86%
5 (A)	30	64	33	66%
6 (A)	31	61	36	63%
7 (C)	17	72	25	74%
8 (C)	18	29	68	30%
9 (C)	19	83	14	85%
10 (C)	19	77	20	79%
11 (C)	20	81	16	83%
12 (C)	24	64	33	66%
13 (C)	24	70	27	72%
14 (C)	24	68	29	70%
15 (C)	24	79	18	82%
16 (C)	24	68	29	70%
17 (C)	24	84	13	86%
18 (C)	24	93	4	96%
19 (C)	28	81	16	83%
20 (C)	30	71	26	73%
Control Group	23	73	24	74%
Autism Group	25	71	26	75%

3.2 Agreement on social referencing

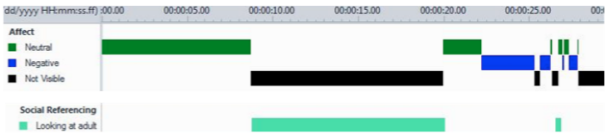
We next measured the agreement on frequency of social referencing, which ranged from 0 to 8 head turns per subject in data from both the human and computer coding. Two-way, average measure, consistency ICC was 0.89 (95% CI of 0.73-0.96), allowing us to infer that the automated analysis has acceptable classification of social referencing. Limitations on exact agreement include the inability of the computer to determine whether a child is looking at a person or an object behind them (which we could add in the future by automatically detecting the other people). We achieved excellent agreement on number of head turns, and may be able to distinguish social referencing from random movement by the coordination of affect and social communication with head turns in deeper analysis of the sequence of events.

3.3 Pattern of behaviors

The next step in developing this measure is to refine the automated method and take into account the sequence of events. A strength of observational measures carried out by experts is that they take into account not only whether a child displays a behavior, but also the sequence and integration of different behaviors. One key way that children with autism are identified is via lack of reciprocal social interactions, which is the act of directing facial expression and communication to others [6, 13]. In order to work toward a measurable difference in autistic children, we must also take into account the temporal pattern of observed events. We propose to measure this by automatically detecting and characterizing the combination of positive affect and social referencing elicited by the iPad-delivered stimulus, or



(a) Control



(b) Autism

Figure 3: Comparison of sequence of behaviors in a control subject and an autistic subject matched on age during the same stimulus. Green is neutral expression, red is positive expression, dark blue is negative expression, black is not visible, and light green is social referencing (looking at adult).

lack thereof. The sequence of automatically coded affective states and social referencing in non-autistic children reveals a pattern of behaviors that may be distinct from that of autistic children, Figure 3. We intend to characterize these patterns and identify group differences and developmental patterns on an expanded group of children.

4. CONCLUSIONS

In this study, we have demonstrated the feasibility of delivering video stimuli in a clinic to young children and collecting usable data on a mobile device. For parents and children who participated in this study, the task was easy and enjoyable. These data suggest that observation of behaviors used to indicate possible risk for autism can be elicited and automatically measured with this app. Our technology has the potential to improve access to autism screening, particularly when evaluation by experts is limited and expensive. Our tool can help to ensure that these evaluations are directed to children who truly have an elevated risk for autism. We interpret these results with the caution that we have not yet demonstrated a group difference on a large sample of children with and without autism. Further refinement of this tool will be necessary before it can be considered a potential automatic screening instrument for autism. However, this is the first step in the development of a new class of tools for automated analyses of child behavior that could be applied more broadly to screening for a range of neurodevelopmental and mental health disorders and monitoring of the development of these disorders and response to treatments over time. A strength of our approach is that we are able to rapidly improve the technology because we are able to collect data on large, heterogeneous samples because the technology is scalable and, thus, enables us to rapidly recruit a normative sample. We hypothesize that children may perform differently in their home or school setting, and we therefore plan to expand data collection into these settings. Broader application and dissemination of tools for assessing early childhood behaviors will improve screening for autism and other neurodevelopmental and mental health symptoms, and early identification and on-going monitoring should improve access to quality interventions that will support child's healthy development and functioning.

5. ACKNOWLEDGMENTS

Funding and support from the Duke Center for Autism and Brain Development, Information Initiative at Duke, Psychiatry Research Incentive and Development Grant, Education and Human Development Incubator Award, Duke Endowment, Duke Clinical and Translational Science Award (TL1TR001116), Duke University School of Medicine Primary Care Leadership Track National Science Foundation, and The Department of Defense. We thank the pediatricians and clinical staff at Duke Primary Care Pediatrics for their support and efforts to improve early screening for autism.

6. REFERENCES

[1] L. Camras, Z. Meng, T. Ujiie, S. Dharamsi, K. Miyake, H. Oster, and J. Campos. Observing emotion in infants: facial expression, body behavior, and rater judgements of responses to an expectancy-violating event. *Emotion*, 2(2):179 – 193, 2002.

[2] G. Dawson, K. Toth, R. Abbott, J. Osterling, J. Munson, A. Estes, and J. Liaw. Early social attention impariments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology*, 40(2):271–283, 2004.

[3] H. Egger and A. Angold. Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Jouranl of Child Psych. and Psychiatry*, 47(3-4):313–337, 2006.

[4] A. Estes, L. Vismara, C. Mercado, A. Fitzpatrick, L. Elder, J. Greenson, and S. Rogers. The impact of parent-devliered intervention on parents of very young children with autism. *Journal of Autism and Developmental Disorders (JADD)*, 44(2):353–365, 2014.

[5] J. Fox and S. Weisberg. *An R companion to applied regression*. Sage, Thousand Oaks, CA, second edition, 2011.

[6] D. Gangi, L. Ibanez, and D. Messinger. Joint attention initiation with and without positive affect: risk group differences and associations with ASD symptions. *Journal of Autism and Developmental Disorders (JADD)*, 44(6):1414–1424, 2014.

[7] K. Hallgren. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1):23–34, 2012.

[8] J. Hashemi, Q. Qiu, and G. Sapiro. Cross-modality pose-invariant facial expression. In *IEEE Conference on Image Processing (ICIP)*, 2015.

[9] J. Hashemi, M. Tepper, T. Spina, A. Esler, V. Morellas, N. Papanikolopoulos, H. Egger, G. Dawson, and G. Sapiro. Computer vision tools for low-cost and non-invasive measurement of autism-related behaviors in infants. *Autism Research and Treatment*, 2014.

[10] D. Huang, M. Storer, F. De la Torre, and H. Bischof. Supervised local subspace learning for continuous head pose estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[11] T. Kanade, J. Cohn, and Y. Tian. Comprehensive

database for facial expression analysis. In *IEEE Conference on Automatic Face and Gesture Recogition (FG)*, pages 46–53, 2000.

[12] J. Larsen, C. Norris, and J. Cacioppo. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40(5):776–785, 2003.

[13] C. Lord, S. Risi, L. Lambrecht, E. Cook, B. Leventhal, P. DiLavore, and M. Rutter. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders (JADD)*, 30(3):205–223, 2000.

[14] D. Matsumoto and H. Hwang. Evidence for training the ability to read microexpressions of emotion. *Motivation and Emotion*, 35(2):181–191, 2011.

[15] J. Nadel and D. Muir. *Emotional development: recent research advances*. Oxford University Press, Oxford, New York, 2005.

[16] Nodlus. *http://www.noldus.com/human-behavior-research/products/the-observer-xt*, 2015.

[17] H. Oster, D. Hegley, and L. Nagel. Adult judgements and fine-grained analysis of infant facial expressions: testing the validaty of a priori coding formulas. *Developmental Psychology*, 28(6):1115–1131, 1992.

[18] S. Ozonoff, A. Iosif, F. Baguio, I. Cook, M. Hill, T. Hutman, R. A. Roger, S., S. Sangha, M. Sigman, M. Steinfeld, and G. Young. A prospective study of the emergence of early behavioral signs of autism a prospective study of the emergence of early behavioral signs of autism. *J Am Acad Child Adolesc Psychiatry*, 49(3):256–266, 2010.

[19] K. Pierce, D. Conant, R. Hazain, R. Stoner, and J. Desmond. Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, 68(1):101–109, 2011.

[20] D. Robins, K. Casagrande, M. Barton, C. Chen, T. Dumont-Mathieu, and D. Fein. Validation of the Modified Checklist for Autism in Toddlers, revised with follow-up M-CHAT-R/F. *PEDIATRICS*, 133(1):37–45, 2014.

[21] O. Rudovic, M. Pantic, and I. Patras. Coupled Gaussian processes for pose-invariant facial expression recognition. *IEEE Trans on Pattern Analysis and Machine Intelligence (PAMI)*, 35(6):1357 – 1369, 2013.

[22] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. Static and dynamic 3D facial expression recognition: a comprehensive survey. *Image and Vision Computing*, 30(10):683–697, 2012.

[23] C. Shan, S. Gong, and P. McOwan. Facial expression recognitino based on local binary patterns: a comprehensive study. In *Image and Vision Computing*, volume 27, pages 803–816, 2009.

[24] H. Tang, M. Hasegawa-Johnson, and T. Huang. Non-frontal view facial expression recognition based on ergodic hidden markov model supervectors. In *IEEE Conference on Mulitmedia and Expo (ICME)*, pages 1202–1207, 2010.

[25] R. C. Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[26] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 532–539, 2013.

[27] L. Yin, X. Wei, J. Wang, and M. Rosato. A 3D facial expression database for facial behavior research. In *IEEE Conference on Automatic Face and Gesture Recogition (FG)*, pages 211–216, 2006.

[28] X. Zhao, E. Dellandréa, and J. Zou. A unified probabilistic framework for automatic 3D facial expression analysis based on a bayesian belief inference and statistical feature models. *Image and Vision Computing*, 31(3):231–245, 2013.