Research in Autism Spectrum Disorders 65 (2019) 14–24

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: www.elsevier.com/locate/rasd

A pilot study to identify autism related traits in spontaneous facial actions using computer vision

Manar D. Samad[a,*], Norou Diawara[b], Jonna L. Bobzien[c], Cora M. Taylor[d], John W. Harrington[e], Khan M. Iftekharuddin[a]

[a] Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA
[b] Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA, USA
[c] Department of Communication Disorder and Special Education, Old Dominion University, Norfolk, VA, USA
[d] Department of Clinical Psychology, Geisinger, Lewisburg, PA, USA
[e] Department of Pediatrics, Children's Hospital of the Kings Daughters, Norfolk, VA, USA

ARTICLE INFO

Keywords:
ASD
Behavioral marker
Differential traits
Facial action units
Computer vision
Spontaneous expressions

ABSTRACT

Background: Individuals with autism spectrum disorders (ASD) may be differentiated from typically developing controls (TDC) based on phenotypic features in spontaneous facial expressions. Computer vision technology can automatically track subtle facial actions to gain quantitative insights into ASD related behavioral abnormalities.
Method: This study proposes a novel psychovisual human-study to elicit spontaneous facial expressions in response to a variety of social and emotional contexts. We introduce a markerless facial motion capture and computer vision methods to track spontaneous and subtle activations of facial muscles. The facial muscle activations are encoded into ten representative facial action units (FAU) to gain quantitative, granular, and contextual insights into the psychophysical development of the participating individuals. Statistical tests are performed to identify differential traits in individuals with ASD after comparing those in a cohort of age-matched TDC individuals.
Results: The proposed framework has revealed significant difference ($p < 0.001$) in the activation of ten FAU and contrasting activations of FAU between the group with ASD and the TDC group. Unlike the TDC group, the group with ASD has shown unusual prevalence of mouth frown (FAU 15) and low correlations in temporal activations of several FAU pairs: 6–12, 10–12, and 10–20. The interpretation of different FAU activations suggests quantitative evidence of expression bluntness, lack of expression mimicry, incongruent reaction to negative emotions in the group with ASD.
Conclusion: Our generalized framework may be used to quantify psychophysical traits in individuals with ASD and replicate in similar studies that require quantitative measurements of behavioral responses.

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by the demonstration of social communication deficits in diagnosed individuals. The deficits in social communications may appear in nonverbal communications, such as abnormal

* Corresponding author at: Tennessee State University, Department of Computer Science, 3500 John A. Merritt Boulevard, PO Box 9604, Nashville, TN, USA.
E-mail address: msama005@odu.edu (M.D. Samad).

https://doi.org/10.1016/j.rasd.2019.05.001
Received 28 September 2018; Received in revised form 4 May 2019; Accepted 8 May 2019
Available online 30 May 2019
1750-9467/

use of facial expressions during social interactions. Facial expressions are primarily a physiological manifestation of facial muscles, which can reflect a variety of mental states, cognitive processes, mimicry of expressions, and other voluntary or involuntary motor actions. Hence, atypical traits in facial expressions may reveal useful behavioral biomarker and differential attributes that promote a better understanding of neurodevelopmental disorders like ASD. Facial expressions have been studied in a number of neurological disorders such as ASD (Stel, van den Heuvel, & Smeets, 2008; Volker, Lopata, Smith, & Thomeer, 2009), Schizophrenia (Hamm, Kohler, Gur, & Verma, 2011), Alzheimer's disease (Seidl, Lueken, Thomann, Kruse, & Schroder, 2012), and Parkinson's disease (Simons, Pasqualini, Reddy, & Wood, 2004). However, these studies involve one or more of several approaches such as subjective encoding of facial expressions, intrusive probing of facial muscle activations, or analyzing 'target' expression imitation skills of TDC. We believe that the recent advancements in facial imaging and computer vision techniques can provide more quantitative, unobtrusive, contextual analysis of facial traits as a promising alternative to current methods in psychophysical studies. In this paper, we propose facial image and computer vision-based novel experimental frameworks that may alleviate a number of methodological limitations and serve as a generalized framework for many of similar applications.

The organization of the remaining paper is as follows. Section 2 provides a background review highlighting the limitations in ASD literature and prospects of using facial imaging in this context. Section 3 outlines the experimental setup, human-study protocol, and the proposed framework for encoding and analyzing facial imaging data. Section 4 presents the results and Section 5 discusses the findings in the context of ASD-related differential traits, and the paper concludes in Section 6.

2. Background review

ASD is a behavioral disorder that relies heavily on measurements of symptoms in behavioral manifestations. However, the precise measurement of behavioral traits is difficult as it typically involves interviews, behavioral observations, and parent-report questionnaires that are subjective in nature (Barbaro & Dissanayake, 2009; Huerta & Lord, 2012; Hirota, So, Kim, Leventhal, & Epstein, 2018). An objective measurement of behavioral symptoms can be an important contribution to ASD research that may improve the reliability of the subjective assessment methods. In this regard, the application of computer vision technology may play a pivotal role in automating and quantifying human’s behavioral expressions and yield higher accuracy than manual scoring. Computational tools may objectively identify the heterogenous characteristics of the disorder leading to an increased understanding of etiology and subsequent interventional planning for ASD. The following section summarizes some of the recent differential studies on facial phenotypes and expressions in individuals with ASD.

2.1. Differential study of facial expressions

ASD-related characteristics have been studied during elicited and posed or imitated facial expressions. A majority of studies use video images to analyze facial expression imitation skills (Guha, Yang, Grossman, & Narayanan, 2016; Volker et al., 2009) or spontaneous facial expressions demonstrated in response to emotional or social stimuli (Adrien et al., 1992; Stel et al., 2008; Yirmiya, Kasari, Sigman, & Mundy, 1989). Stel et al. use videos showing emotions in others to elicit spontaneous facial expressions in participants with ASD (Stel et al., 2008). Participants with ASD are visually evaluated by untrained human raters to identify anomalies in the facial expressions of TDC with ASD. Unfortunately, human raters may miss subtle movement patterns in facial expressions, resulting in subjective data and poor inter-rater reliability. Unlike spontaneous facial expressions, the imitation of expression requires deliberate effort for posing. As such, posed facial expressions may not be representative of the genuine emotional and cognitive processes of participants when exposed to an authentic emotional trigger. Yirmiya et al. qualitatively encode facial expressions in infants and toddlers in a differential study between children with and without ASD (Yirmiya et al., 1989). The facial expressions of children of this age is minimal and can be categorized into only four groups - 'positive', 'negative', 'neutral', 'interest' - in response to human interactions involving toys. Therefore, the study of more complex nonverbal communications during social interactions will require older participants, around the age of adolescence, who are able to communicate verbally, accurately understand and react to social contexts.

ASD researchers have also used facial sensors such as electromyography (EMG) electrodes (Liu, Conn, Sarkar, & Stone, 2008; Rozga, King, Vuduc, & Robins, 2013) or facial markers (Guha et al., 2016) to quantify facial actions in participants with ASD. While adults can adjust to the intrusive application of sensors, children with ASD often suffer from sensory sensitivities, and increased anxiety related to novel experiences (Harrington & Allen, 2014). Hence, the use of EMG has inherent difficulties in this population and may interfere with the ability to collect accurate data.

Apart from intrusiveness, prior EMG studies are limited to probing only two muscles: *Zygomaticus Major* and *Corrugator Supercilli* around the cheek and brow regions, respectively (Liu et al., 2008; Magnée, De Gelder, Van Engeland, & Kemner, 2007; McIntosh, Reichmann-Decker, Winkielman, & Wilbarger, 2006; Oberman, Winkielman, & Ramachandran, 2009; Rozga et al., 2013). The magnitude of activations of these two muscles are studied without clearly connecting the underlying psychophysics of these muscle activations to different stimuli. Furthermore, static images of facial expressions are used to stimulate spontaneous facial responses (McIntosh et al., 2006; Oberman et al., 2009). We believe that dynamic social stimuli are more realistic in the study of differential traits in facial expressions. In this paper, we propose dynamic audio-visual stimuli containing an array of emotional contexts as seen during social encounters. We present a generalized framework that may be followed in a diverse set of psychophysical studies on spontaneous facial expression analysis.

2.2. Computer vision and ASD research

The methodological constraints in the ASD literature (Section 2.1) may be overcome by using advanced facial imaging and computer vision techniques. Facial imaging data are extensively used in a wide range of applications from surveillance and biometrics, to human computer interactions (HCI). Jaiswal et al. have used depth sensors (e.g., Kinect) to leverage dynamic 3D or 4D facial data in classifying individuals with ASD (Jaiswal, Valstar, Gillott, & Daley, 2017). Despite advances in facial imaging and algorithms for automatic tracking of facial features (Samad, Bobzien, Harrington, & Iftekharuddin, 2016; Zhao, Dellandrea, Zou, & Chen, 2013), there is a dearth of computational imaging studies in the field of psychology and behavioral science, especially to increase our understanding of complex neurodevelopmental disorders like ASD. Recently, computer vision and imaging techniques have been used in a number of studies related to ASD research. Campbell et al. have studied the attention to 'name calls' by tracking head motions from videos of toddlers with ASD (Campbell et al., 2018). Egger et al. have studied the feasibility of an i-phone based computer vision application by classifying facial expressions of toddlers into 'positive' and 'negative' groups (Egger et al., 2018). Their stimuli include videos of bunny and bubbles to investigate the association between facial expressions and ASD risk. Similar method is used to classify facial expressions of individuals with ASD into 'happy' versus 'other' (Hashemi et al., 2018). The facial expression imitation skills are evaluated in individuals with ASD by using computer vision application that recognizes six basic emotional expressions from facial images (Leo et al., 2015). Individuals with ASD are instructed and even corrected while imitating four expressions: happiness, sadness, fear, and anger (Leo et al., 2018). The intensity of constituent facial action units of these imitated expressions is analyzed to determine the accuracy of their imitation skill.

Computer vision technology can be used to assess more complex differential traits when individual with ASD in adolescence encounter complex social scenarios. The analysis of complex and subtle facial expressions in terms of an array of facial action units (FAU), defined by Facial Action Coding System (FACS) (Ekman & Friesen, 1978) may provide more granular and insightful data for this investigation. We hypothesize that computer vision technology can provide more granular stratification of spontaneous facial expressions in terms of FAU intensities that may objectively determine anomalies in social and behavioral expressions of individuals with ASD.

2.3. Contributions

This paper proposes a novel framework to promote quantitative, granular, and contextual analysis of facial expressions for differential studies on ASD. This work delves into complex social expressions of school-aged children with ASD to go beyond the study of facial expression 'imitation skills' or the classification binary emotional expressions (positive versus negative) in ASD literature. A psychovisual experiment is performed to elicit spontaneous facial expressions using dynamic social stimuli and simultaneously track using a markerless facial motion capture system. Spontaneous response of ten facial locations are represented in ten representative facial action units, which yield more granular data compared to using two EMG electrodes on the face. The activation of ten facial actions is also studied in response to varying emotional contexts to gain useful insights into the psychophysical development of individuals with ASD. To our knowledge, this is one of the first studies to use computer vision tracking of spontaneous facial action units to quantify ASD related differential traits in school-aged children.

3. Experimental design and analysis

The protocols for collecting and analyzing facial data from participating human participants are approved by both Institutional Review Boards (IRBs) at Eastern Virginia Medical School (EVMS) and Old Dominion University. The sections below describe the experimental design and protocol for the proposed study.

3.1. Participants

Twenty school-aged children, between the ages of eight to seventeen years, are recruited for this study. Ten participants (ten males) are diagnosed with ASD, and ten (seven males) are typically developing controls (TDC). Each individual in the group with ASD has received a confirmed diagnosis of ASD by clinicians at EVMS using the autism diagnostic observation schedule (ADOS) (Lord et al., 2000). According to the Center for Disease Control and Prevention (CDC) prevalence data, the majority (69%) of the children diagnosed with ASD are classified as having intellectual ability with intelligence quotient (IQ) scores over 85 (44%) or as having borderline intellectual ability (25%) with IQ scores between 71 and 85 (Baio et al., 2018). The proposed task in this study requires each participant to view and understand the social stimuli, hence we exclude individuals with an IQ less than 70 to ensure the intellectual ability required for the given task. The average age of the group members with ASD is 13.5 $\pm$ 2.37 years. The average age of the TDC group members is 13.1 $\pm$ 3.31 years. Mean ages between the two groups are not significantly different ($t(18) = 0.4$, $p = 0.69$). Prior to enrollment, a candidate participant is screened over the phone following inclusion and exclusion criteria as per the study protocol. Once a participant is found eligible, the participant and the parent of the participant complete the informed consent and assent processes, which briefly outline the goals and tasks in the study.

3.2. Stimuli

The purpose of the social stimuli is to elicit spontaneous facial expressions, which may in turn reflect underlying cognitive processes and perception of stimuli. Audio-visual stimuli, involving a computer simulated avatar character, are used in prior studies for ASD therapeutic intervention (Ploog, Scharf, Nelson, & Brooks, 2013). Individuals with ASD have been found to be more engaged with Avatar characters and computer simulated environments than during interactions with real people (Moore, McGrath, & Thrope, 2000). The application of avatar-based intervention in the ASD literature has been promising (Hopkins et al., 2011; Ploog et al., 2013), which motivates the use of avatar-based stimuli to elicit spontaneous facial responses in this study. The persona of a real human actor is used to create an animated avatar character. The avatar character imitates the actor's speech and facial actions in a story-telling scenario, talking about experiences related to his school-life and social engagement. The content and context of the story incorporate a range of emotional expressions that are appropriate for the developmental level of school-age group. The speech and facial animation of the avatar character are recorded as a four-minute video to be used as the audio-visual stimuli.

3.3. Acquisition of facial data

We use a commercially acquired real-time facial motion capture system, known as *faceshift* (www.faceshift.com, version 2015.2) studio is a facial motion tracking and animation application primarily developed for rendering 4D animated Avatar faces from human actors. The application is currently acquired by Apple Inc. and the technology is also used in animations of Star Wars movies. *faceshift* studio uses *PrimeSense* to acquire real-time and dynamic 3D facial data for tracking facial actions. First, a synthetic deformable 3D facial model is calibrated by the neutral face of the participating individual or actor. The system is then used to track distinctive facial muscle movements of the participant, which are simultaneously mapped on the calibrated 3D facial model. In this process, facial muscle activations are encoded into an array of distinctive facial actions known as blendshapes. Previously, these blendshapes have been manually evaluated for tracking sensitivity and subsequently used as facial action units to differentiate patients with Schizophrenia (Tron, Peled, Grinsphoon, & Weinshall, 2016). In this study, *faceshift* studio serves two major applications: 1) in rendering avatar-based audio-visual stimuli and 2) in acquiring spontaneous facial action data of participants in response to the stimuli (Fig. 1). Note that *faceshift* tracking of facial motion is robust to occlusions such as beard, glasses as shown in Fig. 2(a) and is also reliable for a wide range of facial poses.

3.4. Tasks and procedures

The participant is seated in an adjustable chair facing the 68″ Television (TV) display as shown in Fig. 2(b). The TV is used to display the audio-visual stimuli presented by the Avatar character. The *PrimeSense* sensor is adjusted on a tripod and positioned in a way to avoid interfering with the participant's line-of-sight. The task of the participants is to visualize, listen, and understand the four-minute story content narrated by the animated Avatar as shown in Fig. 2(c). The onset of *faceshift* studio for facial data collection, the initiation of the audio-visual stimuli, and the termination of data collection are performed automatically to allow time synchronization between the stimuli and response data.

3.5. Analysis of facial data

The following sections summarize the analytical framework for processing temporal facial activation data obtained from the *faceshift* system.

3.5.1. Analysis of facial action data

The *faceshift* studio tracks and records the magnitude of different distinctive facial actions. The time-varying activation data are sampled at one sample per second and the magnitude of activation is normalized between the values of 0 and 1. For subsequent analysis, we select ten representative facial actions and annotate them by facial action units (FAUs) as shown in Fig. 3. These ten FAUs are

chosen to cover at least six prototypical emotional expressions (happy, sad, fear, anger, surprise, disgust) in humans (Bartlett, Hager, Ekman, & Sejnowski, 1999). Table 1 summarizes contributions of ten proposed FAUs to these prototypical emotional expressions. More flexibly, FAUs can represent subtle facial responses that may not be broadly categorized into one of the prototypical emotional expressions. Therefore, the magnitude of each FAU is analyzed to study the effect of stimuli contexts, FAU locations, and study groups on facial activation.

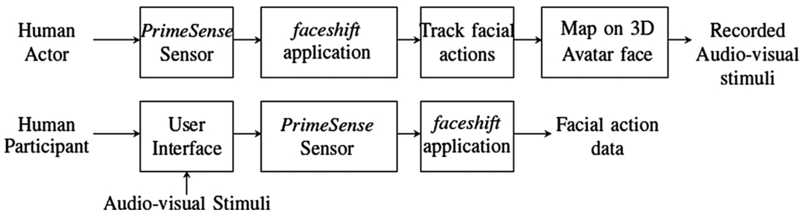

Fig. 1. Application of *faceshift* facial motion capture system. Top: rendering of avatar-based audio-visual stimuli mapping the persona of an actor. Bottom: Collecting facial activation data of participants in response to audio-visual stimuli.

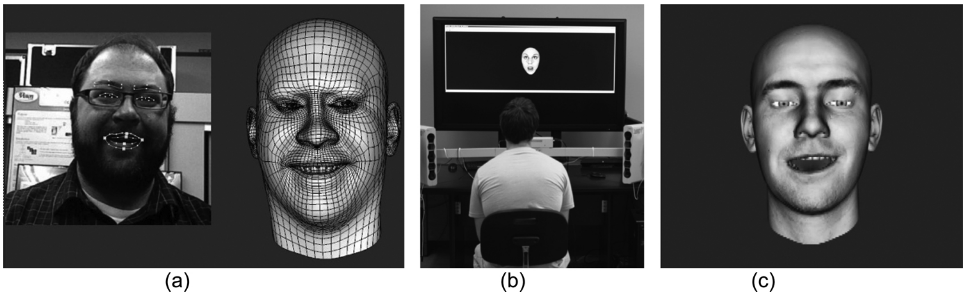

Fig. 2. Experimental setup. (a) Real-time, markerless, and robust tracking of facial key points of the actor and mapping of facial motion on a 3D deformable facial model; (b) A participant is seated in front of the 68″ TV; (c) The animated avatar's face used for narrating the story as the audio-visual stimulus.

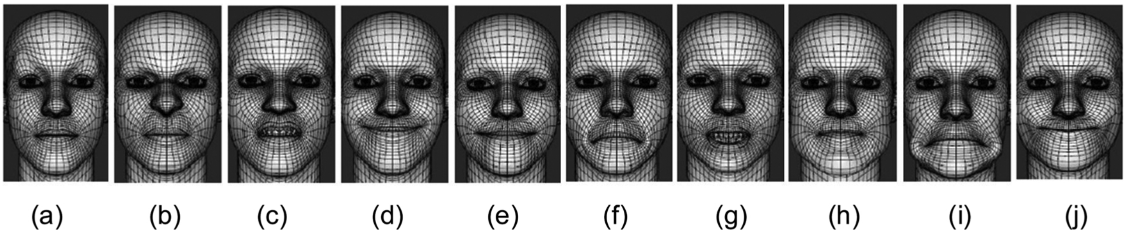

Fig. 3. Synthetically generated ten facial action units (FAU) on a 3D deformable facial model for visualization using *faceshift* studio. (a) FAU 1, (b) FAU 6, (c) FAU 10, (d) FAU 12, (e) FAU 14, (f) FAU 15, (g) FAU 16, (h) FAU 17, (i) FAU 20, (*j*) FAU 24.

Table 1
FACS-based encoding of facial actions tracked by *faceshift* facial motion capture system.

Facial Actions in *faceshift*	Facial Actions Units	Facial Actions Facial Action FAU present in Emotional Expressions
Brows up	AU 1	Sadness, Surprise Fear
Cheek raise	AU 6	Happiness, Scorn
Upper lips	AU 10	Disgust
Smile	AU 12	Happiness
Dimple	AU 14	Depression
Mouth frown	AU 15	Disgust, Sadness
Lower lip down	AU 16	Disgust
Chin raise	AU 17	Disgust
Lips stretch	AU 20	Fear

3.5.2. Data description

We organize the acquired facial response data in two data formats. The data format one (DF1) included magnitude of facial response (y) from each of ten facial locations (10 FAUs) at 240 time points (in response to 240-second stimuli) for each participant belonging to one of the two groups (ASD, TDC). The size of DF1 after including ten participants for each group is 2 × 10 (participants) X10 (facial locations) X 240 (time points). The data format two (DF2) summarizes DF1 and has 10 stimuli points after dividing 240-second stimuli into 10 contexts (Table 2). The magnitude of facial response (y) is averaged within the period of each context to

Table 2
Temporal breakdown of the proposed four-minute audio-visual stimuli in terms of social contexts and emotional contents.

Stimuli index	Context of stimuli	Total time (s)	Avatar emotion
1	Introduce school and friends	40	Neutral
2	Plan for watching movie with friends	14	Excited
3	Pizza and anchovy in pizza	16	Disgusted
4	Tony took my notebook and forgot to return	29	Angry
5	Missing notebook, test tomorrow!	28	Anxious & Sad
6	Tony shows up with notebook and a candy bar	13	Surprise
7	In the exam hall with the test	25	Tensed
8	Saturday bowling party	12	Happy
9	Bowling performance	42	Suspense
10	Saying good bye	7	Neutral

study the effect of different social contexts on facial response (y). The size of DF2 is therefore 2 × 10 (participants) X 10 facial locations X 10 stimuli points. The response data in DF1 are used to obtain the correlation coefficient between the time-sampled activations of FAUs within each group. A prototypical facial expression (e.g. happy, anger, surprise) may involve concurrent activations of multiple FAUs. We hypothesize that TDC will group show more concurrent and correlated activations of FAUs compared to the group with ASD. Additionally, we perform several statistical tests on both data formats.

3.5.3. Statistical analyses

We use generalized linear models (GLM) in the mixed effect context to understand the magnitude of facial activation. In this model, the group (ASD vs TDC) is included as the fixed effect. Audio-visual stimuli and facial locations are included in the random effect model. The random noise associated with the errors is assumed to have normal distribution with zero mean and σ_ε^2 variance. The covariance used for the random components is of unstructured covariance structure (type CS). The model equation is represented as follows.

$$Y = X\beta + Zu + \varepsilon$$

Here, Y is the response of 20 participants at 240 time periods (for DF1), which accounts for the groups, stimuli, and facial locations. The stimuli and facial locations are assumed to be participant specific random effects. Here, X is the 240 × 20 × 2 dimensional design matrix, β is the 2-dimensional co-efficient vector of fixed effect of group parameters. The random effect matrix Z is of 240 × 20Xq dimension and u is the q-dimensional (q = 20) coefficient vector of random effect parameters, per participant, with some correlation between the random effects. Our random effect model includes ten facial locations, 240 stimuli, and ten participants within each group. For each of these ten facial locations, we assume the difference between the groups, if significant, including participants (q = 20; total participants = 20). In this model, ε is the vector of residual errors. We summarize a total of 240 × 20 model parameter estimates per group plus the variance of the residual error for the full mixed effect model. The mixed model is obtained using SAS procedure *code proc mixed*. The significance level is chosen at α = 0.05 for all cases. We also report the p-values associated with the correlation coefficients between time-sampled activations of FAUs.

4. Results

This section presents results following the analysis of spontaneous facial action data as outlined in the previous section.

4.1. Visualization of spontaneous facial expressions data

The magnitudes of FAU activations are averaged across ten participants for each group to show the activation heat maps in Fig. 4. Facial activation data are interpolated on a 50-by-50 color grid to visualize activation of ten FAUs in response to ten stimuli contexts (SC). The comparison of facial activation maps between the group with and without ASD reveals several interesting traits. As expected, the mean activations of FAUs are higher in the TDC group than those in the group with ASD. The TDC group shows high activation of FAU 12 (smile expression) at the beginning of the stimuli, which drops in magnitude when the avatar shows anxiety and sadness (SC 5) and tensed expressions (SC 7). The smile expression peaks again when the avatar expresses a pleasant surprise (SC 6) and happiness (SC 8) (See Fig. 4(a) and Table 2). These results demonstrate that spontaneous facial responses in the TDC group are congruent to stimuli contexts, which is an evidence of emotional empathy and mimicry.

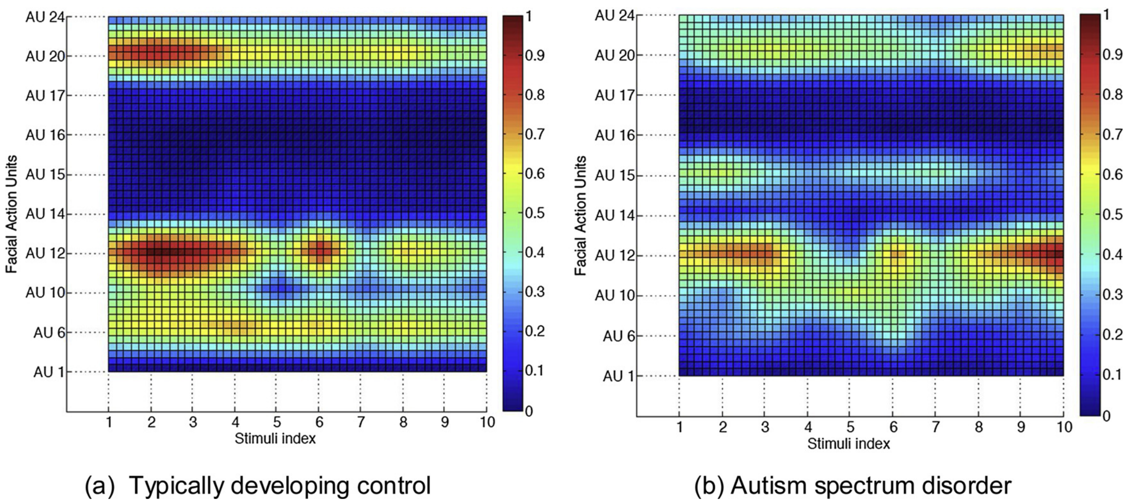

Fig. 4. Visualization of facial activation maps for ten different facial action units in response to ten different stimuli contexts (stimuli index). The magnitudes of facial actions are averaged over all ten participants for each group.

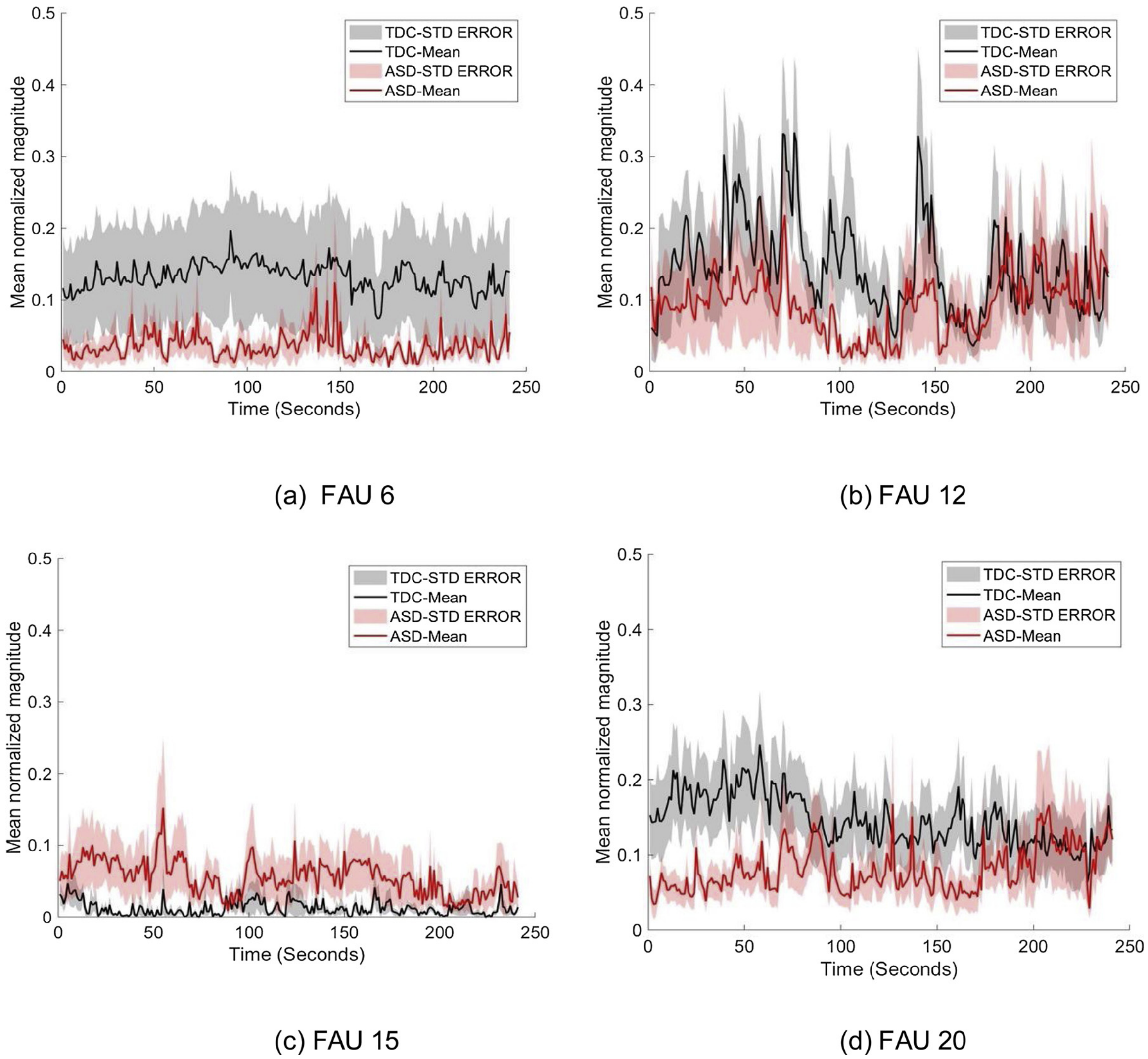

Fig. 5. Temporal activation of four FAUs averaged across ten participants in each group in response to four-minute audio-visual stimuli. a) FAU 6, b) FAU 12, c) FAU 15, and d) FAU 20. Shaded regions represent standard error (STD ERROR).

In contrast, the group with ASD shows lower activation of FAU 12 and almost an absence of FAU 6 (See Fig. 4(b)). Only at the event of the Avatar's pleasant surprise (SC 6), does the group with ASD respond with equal activations for FAU 12 and FAU 6. Unusual activations of mouth frown (FAU 15) have been detected in the group with ASD in response to the Avatar's excitement (SC 2), disgust (SC 3), and tensed (SC 7) expressions. The upper lip is slightly raised (FAU 10) while seeing the avatar anxious (SC 5). The lip stretch action (FAU 20) is minimal in the group with ASD compared to the TDC group. Unlike the TDC group, the group with ASD shows traces of smile expression (FAU 12) when the avatar narrates the bowling game with suspense (SC 9) and leaves the scene with a neutral expression at the end of the stimuli (SC 10). Unusual activations of FAU and absence of appropriate FAUs in response to different emotional contexts may suggest incoherence or lack of empathy in individuals with ASD. Fig. 4 shows that the activations of FAUs 1, 14, 16, 17, and 24 are relatively lower than the other five FAUs. Hence, the five most active FAUs (6, 10, 12, 15, 20) are used in subsequent analyses. The temporal activation of FAU clearly reveals in Fig. 5(a, b, d) that the TDC group yields higher mean activations of FAU 6 (cheek raise), FAU 12 (smile), and FAU 20 (lip stretch), respectively than those of the group with ASD. Interestingly, the group with ASD yields a higher mean activation of FAU 15 (mouth frown) than that of the TDC group (Figs. 4(b) 5(c)).

4.2. Summary of mixed effect models

Fixed effects models using DF1 reveals that both groups ($p < 0.0001$) and facial locations ($p < 0.0001$) have significant effects on time-sampled facial response (y). That is, the TDC group has yielded significantly higher magnitude of facial response than the group with ASD. The ten facial locations (ten FAUs) have resulted in significantly different magnitude of activation (y) within the

Table 3
Tukey grouping of facial action units (FAU) using Tukey's studentized range (HSD) test for magnitude of facial response. Within each group, mean activations (MA) with the same letter are not statistically significant.

TDC										
FAU	20	12	6	10	24	17	14	15	1	16
MA	0.146	0.145	0.132	0.070	0.033	0.020	0.017	0.012	0.006	0.004
Grouping	A	A	B	C	D	E	E, F	F, G	G, H	H
ASD										
FAU	12	20	10	15	24	6	14	17	1	16
MA	0.094	0.080	0.063	0.056	0.046	0.036	0.020	0.012	0.009	0.005
Grouping	A	B	C	C	D	E	F	F, G	G	G

group with ASD ($p < 0.0001$), within the TDC group ($p < 0.0001$), and overall for the groups together ($p < 0.0001$). We further perform Tukey's HSD tests to investigate pair-wise difference in the activation of ten FAUs separately within the group with ASD and the TDC group. Table 3 shows clustering of ten FAUs within each group in terms of statistical significance. The top five most active FAUs account for 80% and 90% of total mean activation for the group with ASD and the TDC group, respectively. In contrast to the TDC group, the group with ASD has shown significant difference between FAU 12 and FAU 20 but no difference between FAU 10 and FAU 15.

The GLM above additionally reveals that the assumptions of fixed effects model are not met since the normality assumption is not satisfied based on the Q-Q plot and equality of variance assumption is not met, which suggests a heterogeneity in the dataset. These limitations can be overcome by considering random and mixed effects models. Furthermore, there is correlation among the responses and the residuals within the same participant. The random effect model can capture the relative similarity of observations in the same participant. Therefore, this study proposes mixed effects model (also known as hierarchical linear or multilevel model) as a flexible class of model that can handle such correlation or dependence structure in data. The generalized estimating equation (GEE) model used in this study yields GEE coefficients to represent expected differences within a population, whereas the coefficients in random effects model represent differences within an individual.

In this study, we include randomness in facial locations for the mixed effect models. Since FAU 16 yields the lowest magnitude of activation for both groups (Table 3), we use FAU 16 as a reference to contrast the estimates of other FAU activations in the mixed model. The mixed model is conversant and the goodness of fit in terms of -2log likelihood is satisfactorily small for the TDC (-33974.2) and the group with ASD (-45253.6), with estimated associated variances of 0.00013 and 0.000082, respectively. Table 4 shows estimates and p-values for FAU activations in reference to FAU 16 within each group. Almost all FAUs yield significantly higher activations in reference to the baseline FAU 16 and individual effects are evident in t-values of the tables. The TDC group and the group with ASD show high t-values for FAUs 6, 10, 12, 20 and FAUs 10, 12, 15, 20, 24, respectively. The DF2 has yielded similar results as DF1 with significant difference in FAU ($p < 0.0001$) and groups ($p = 0.0005$), however, the ten stimuli contexts are not significantly different in producing facial response ($p = 0.68$). Confidence intervals in Table 5 show that most activations of FAU due to stimuli context do not include zeros. Hence, FAU activations due to any stimuli context is significantly different from zero. The rank ordering of the stimuli contexts based on response (y) suggests contrasting activations at stimuli contexts 4 and 10 for the two groups.

Table 4
Mixed effect model estimates with standard error 0.003.

	TDC					ASD				
FAU	Estimate	t value	Pr > |t|	95% Confidence Interval		Estimate	t value	Pr > |t|	95% Confidence Interval	
				Lower	Upper				Lower	Upper
1	0.002	0.67	0.5034	0.004	0.009	0.004	1.41	0.1578	−0.001	0.009
6	0.129	37.29	< .0001	0.122	0.135	0.030	11.11	< .0001	0.025	0.036
10	0.066	19.17	< .0001	0.060	0.073	0.058	21.20	< .0001	0.052	0.063
12	0.141	40.89	< .0001	0.134	0.148	0.089	32.57	< .0001	0.083	0.094
14	0.014	3.92	< .0001	0.007	0.020	0.014	5.25	< .0001	0.009	0.020
15	0.009	2.55	0.0109	0.002	0.016	0.051	18.72	< .0001	0.046	0.056
17	0.016	4.59	< .0001	0.009	0.023	0.007	2.66	0.0078	0.002	0.013
20	0.142	41.30	< .0001	0.136	0.149	0.075	27.62	< .0001	0.070	0.080
24	0.029	8.40	< .0001	0.022	0.036	0.041	15.09	< .0001	0.036	0.046
16	0	.	.	.	.	0	.	.		

Table 5
Response of different stimuli contexts (Table 2) in decreasing order of mean magnitude of spontaneous facial activations for the TDC group and the group with ASD.

TDC				ASD			
Stimuli	Mean	95% Confidence Limits		Stimuli	Mean	95% Confidence Limits	
2	0.077	0.053	0.101	10	0.050	0.033	0.067
3	0.073	0.049	0.097	3	0.048	0.031	0.065
4	0.066	0.042	0.090	6	0.047	0.030	0.064
6	0.064	0.040	0.088	2	0.046	0.029	0.063
1	0.064	0.040	0.088	9	0.044	0.027	0.061
8	0.057	0.033	0.081	1	0.043	0.026	0.060
5	0.052	0.028	0.076	8	0.041	0.024	0.058
7	0.049	0.025	0.073	5	0.039	0.022	0.056
9	0.048	0.024	0.072	4	0.039	0.022	0.056
10	0.045	0.021	0.069	7	0.035	0.018	0.052

4.3. Correlation in facial action unit activation

The cross-correlation between time-sampled FAU activations shows the concurrency in these activations. The five active FAUs (6, 10, 12, 15, 20) form a total of ten pairs of FAUs. Table 6 shows cross-correlation between FAU pairs and corresponding p-values. Four FAU pairs (6–12, 10–12, 10–20, and 12–20) show correlation above or equal to 0.5 with $p < 0.001$ in the TDC group. The highest correlation co-efficient is achieved by the FAU 10–20 pairs (0.76, $p < 0.001$) in the TD group, which represents simultaneous activations of upper lip and lips stretch (Table 1). None of the FAU pairs appear correlated in the group with ASD except for the pair of FAU 15 and FAU 20. This pair of FAUs reveals a negative correlation of -0.56 ($p < 0.001$), which is not significant in the TDC group (-0.11, $p = 0.08$). The lack of correlation or concurrency in FAU activations may be a useful quantitative measure to evaluate ASD related differential traits in spontaneous facial expressions.

5. Discussion

This study proposes a novel framework to investigate ASD-related differential traits in spontaneous facial expressions using advanced computer vision techniques. The findings of this study can be summarized as follows. First, the magnitude of FAU activations for the group with ASD is observed to be significantly lower than that of the TDC group. This proof-of-concept demonstrates an unobtrusive way to quantify bluntness in spontaneous facial expressions of individuals with ASD. Second, despite traces of bluntness in general, the group with ASD demonstrates higher activations of FAU 15 (mouth frown) when compared to the TDC group. Similarly, Jaiswal et al. have shown FAU 15 to be one of the most discriminating action units for classifying between control and condition (with ASD or ADHD) groups (Jaiswal et al., 2017). Hence, mouth frown (FAU 15) may be a useful target to objectively measure ASD related traits in facial expressions. Third, spontaneous activations of different FAUs are studied in response to different stimuli and social contexts. The TDC group shows mimicry and empathy by regulating their facial expressions in response to social contexts and emotional contents of the displayed stimuli (See Section 4.1). However, the group with ASD shows limited activation of FAUs in response to particularly encountering negative emotional states such as fear, anxiety, and sadness. Fourth, the TDC group shows correlated and concurrent activations of several FAU pairs: 6–12, 10–12, 10–12, which are found to be absent in the group with ASD. The lack of concurrence in the dynamic activation of several FAU pairs suggests a promising differential marker for the group with ASD. For example, the group with ASD shows flatness in the activation of FAU 6 relative to the activations of FAU 12 (Figs. 4(a), 5 (a, b)). FAU 6 is one of the two major constituents (FAUs 6 and 12) of the smile expression. Both FAUs are almost equally activated in spontaneous facial expressions of the TDC group. The flatness in the activation of FAU 6, high negative correlation

Table 6
Correlation coefficient between facial activations in response to four-minute visual stimuli. The upper and lower tringles of the diagonal represent results of the TDC group and those for the group with ASD, respectively.

ASD/TDC	AU - 6	AU - 10	AU - 12	AU - 15	AU – 20
AU – 6	–	0.33 ($p = 0.00$)	0.51 ($p = 0.00$)	0.08 ($p = 0.22$)	0.19 ($p = 0.003$)
AU – 10	0.14 ($p = 0.03$)	–	0.76 ($p = 0.00$)	−0.26 ($p = 0.00$)	0.50 ($p = 0.00$)
AU – 12	0.14 ($p = 0.03$)	−0.11 ($p = 0.05$)	–	−0.13 ($p = 0.04$)	0.53 ($p = 0.00$)
AU – 15	0.03 ($p = 0.67$)	−0.12 ($p = 0.09$)	−0.22 ($p = 0.0005$)	–	−0.11 ($p = 0.08$)
AU – 20	0.10 ($p = 0.14$)	0.03 ($p = 0.66$)	0.44 ($p = 0.00$)	−0.56 ($p = 0.00$)	–

between FAU 15 and FAU 20, unusual but consistent manifestations of FAU 15 – all are candidates for differential traits of the group with ASD. These atypical responses in the face may be linked to the pathophysiology of the neurodevelopmental disorder and may be used as behavioral biomarkers related to facial expressions in individuals with ASD.

These quantitative data serve as a proof-of-concepts, provide a quantitative means to objectively identify deficits in processing emotional and social contexts (Rump, Giovannelli, Minshew, & Strauss, 2009) and, thus, validate our proposed framework for quantitative assessment of atypical traits in spontaneous facial response. This paper shows feasibility of a computer vision framework that can be used to perform quantitative and fine-grain analysis of facial expressions in response to a variety of custom stimuli to better understand and identify differential traits of individuals with ASD.

5.1. Sample size and statistical power

The sample size of participants per study group may appear small compared to large studies. Our experiment involves repeated measurements of each participant at different time points and facial locations, which results in much smaller error for the variability in time between and within participants. The power calculation based on 2000 samples (10 stimuli X 10 facial locations X10 participants X 2 groups) and the variances associated with the mixed effect models (around 0.01 or 0.02 for each group) lead to a power close to one. Such power calculation is based on the method proposed by Green and MacLeod (Green & MacLeod, 2016) for mixed model where power calculation is done once the model and effect size have been specified. Furthermore, repeated measures design increases statistical power for detecting changes in facial activations when compared to traditional cross-sectional design of experiments as conceptualized in (Guo, Logan, Glueck, & Muller, 2013; Oberfeld & Franke, 2013). The multiplicity is accounted in the sense that the mixed model accounts for correlations and repeated measures.

5.2. Limitations

Despite promising results and successful demonstration of a computer vision framework in ASD research, this study has a number of limitations. The study has excluded participants with ASD who demonstrated intellectual disability (IQ < 70), which may limit the generalizability of the results to all children with ASD. We believe that the comprehension of the social stimuli in the given tasks warrants this exclusion. As (Mackenzie & Wonders, 2016) have asserted, it is important to include all individuals with disabilities (i.e. ADHD) regardless of IQ score; however, it may be necessary to utilize IQ parameters when the research tasks are too difficult for an individual with an intellectual disability to understand. The participants with ASD have a confirmed diagnosis of ASD using ADOS, however, a more recent ADOS outcomes may be useful for this research. However, our study provides suitable first step to determining ASD related characteristics in the repeated values of the change in spontaneous facial activations associated with stimuli and locations.

6. Conclusions and future work

This work proposes real-time capture of spontaneous facial expressions to study ASD-related differential traits in the activation of subtle facial action units. The fine-grain analysis of facial action units shows quantitative ways to identify a number of differential traits in spontaneous response of individuals with ASD. These findings serve as a proof-of-concepts and show the feasibility and effectiveness of our proposed experimental framework in detecting behavioral biomarkers for ASD. In comparison to the current practice of qualitative evaluation, the proposed framework may be utilized to quantify participant-specific impairments in social engagement and responses. In the future, scopes for these computational frameworks need to be extended for a larger population to develop adaptive systems for participant-specific screening and treatment planning of individuals with ASD.

Acknowledgements

This work has been partially funded by a grant from Old Dominion University Office of Research Multidisciplinary Funding Program (#300261-010) and a grant from NSF (ECCS 1310353).

References

Adrien, J. L., Perrot, A., Sauvage, D., Leddet, I., Larmande, C., Hameury, L., ... Barthelemy, C. (1992). Early symptoms in autism from family home movies. Evaluation and comparison between 1st and 2nd year of life using I.B.S.E. scale. *Acta Paedopsychiatrica, 55*(2), 71–75.

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., ... Dowling, N. F. (2018). Prevalence of autism Spectrum disorder among children aged 8 years — Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveillance Summaries, 67*(6), 1–23.

Barbaro, J., & Dissanayake, C. (2009). Autism Spectrum disorders in infancy and toddlerhood: A review of the evidence on early signs, early identification tools, and early diagnosis. *Journal of Developmental & Behavioral Pediatrics, 30*(5), 447–459.

Bartlett, M. S., Hager, J. C., Ekman, P., & Sejnowski, T. J. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology, 36*(2), 253–263.

Campbell, K., Carpenter, K. L. H., Hashemi, J., Espinosa, S., Marsan, S., Borg, J. S., ... Dawson, G. (2018). Computer vision analysis captures atypical attention in toddlers with autism. *Autism*.

Egger, H. L., Dawson, G., Hashemi, J., Carpenter, K. L. H., Espinosa, S., Campbell, K., ... Sapiro, G. (2018). Automatic emotion and attention analysis of young children at home: A ResearchKit autism feasibility study. *Npj Digital Medicine*.

Ekman, P., & Friesen, W. (1978). *Facial action coding system (FACS): A technique for the measurement of facial action.* Palo Alto, CA: Consulting.

Green, P., & MacLeod, C. J. (2016). SIMR : An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498.

Guha, T., Yang, Z., Grossman, R. B., & Narayanan, S. S. (2016). A computational study of expressive facial dynamics in children with autism. *IEEE Transactions on Affective Computing,* 1.

Guo, Y., Logan, H. L., Glueck, D. H., & Muller, K. E. (2013). Selecting a sample size for studies with repeated measures. *BMC Medical Research Methodology, 13*, 100.

Hamm, J., Kohler, C. G., Gur, R. C., & Verma, R. (2011). Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods, 200*(2), 237–256.

Harrington, J. W., & Allen, K. (2014). The clinician's guide to autism. *Pediatrics in Review, 35*(2), 62–78.

Hashemi, J., Dawson, G., Carpenter, K. L. H., Campbell, K., Qiu, Q., Espinosa, S., ... Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing* 1–1.

Hirota, T., So, R., Kim, Y. S., Leventhal, B., & Epstein, R. A. (2018). A systematic review of screening tools in non-young children and adults for autism spectrum disorder. *Research in Developmental Disabilities, 80*, 1–12.

Hopkins, I. M., Gower, M. W., Perez, T. A., Smith, D. S., Amthor, F. R., Wimsatt, F. C., ... Biasini, F. J. (2011). Avatar assistant: Improving social skills in students with an ASD through a computer-based intervention. *Journal of Autism and Developmental Disorders, 41*(11), 1543–1555.

Huerta, M., & Lord, C. (2012). Diagnostic evaluation of autism spectrum disorders. *Pediatric Clinics of North America, 59*(1), 103–111 xi.

Jaiswal, S., Valstar, M. F., Gillott, A., & Daley, D. (2017). Automatic detection of ADHD and ASD from expressive behaviour in RGBD data. *Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition,* 762–769. https://doi.org/10.1109/FG.2017.95.

Leo, M., Carcagnì, P., Distante, C., Spagnolo, P., Mazzeo, P., Rosato, A., ... Lecciso, F. (2018). Computational assessment of facial expression production in ASD children. *Sensors, 18*(11), 3993.

Leo, M., Del Coco, M., Carcagni, P., Distante, C., Bernava, M., Pioggia, G., ... Palestra, G. (2015). *Automatic emotion recognition in robot-children interaction for ASD treatment. 2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*537–545.

Liu, C., Conn, K., Sarkar, N., & Stone, W. (2008). Physiology-based affect recognition for computer-assisted intervention of children with Autism Spectrum disorder. *International Journal of Human-computer Studies, 66*(9), 662–677.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). 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, 30*(3), 205–223.

Mackenzie, G. B., & Wonders, E. (2016). Rethinking intelligence quotient exclusion criteria practices in the study of attention deficit hyperactivity disorder. *Frontiers in Psychology, 7*, 794.

Magnée, M. J. C. M., De Gelder, B., Van Engeland, H., & Kemner, C. (2007). Facial electromyographic responses to emotional information from faces and voices in individuals with pervasive developmental disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 48*(11), 1122–1130.

McIntosh, D. N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J. L. (2006). When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. *Developmental Science, 9*(3), 295–302.

Moore, D. J., McGrath, P., & Thrope, J. (2000). Computer aided learning for people with autism – a framework for research and development. *Innovation in Education and Training International, 37*(3), 218–228.

Oberfeld, D., & Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data. *Behavior Research Methods, 45*(3), 792–812.

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2009). Slow echo: Facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders. *Developmental Science, 12*(4), 510–520.

Ploog, B. O., Scharf, A., Nelson, D., & Brooks, P. J. (2013). Use of computer-assisted technologies (CAT) to enhance social, communicative, and language development in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(2), 301–322.

Rozga, A., King, T. Z., Vuduc, R. W., & Robins, D. L. (2013). Undifferentiated facial electromyography responses to dynamic, audio-visual emotion displays in individuals with autism spectrum disorders. *Developmental Science, 16*(4), 499–514.

Rump, K. M., Giovannelli, J. L., Minshew, N. J., & Strauss, M. S. (2009). The development of emotion recognition in individuals with autism. *Child Development, 80*(5), 1434–1447.

Samad, M. D., Bobzien, J. L., Harrington, J. W., & Iftekharuddin, K. M. (2016). Non-intrusive optical imaging of face to probe physiological traits in Autism Spectrum disorder. *Optics and Laser Technology, 77*.

Seidl, U., Lueken, U., Thomann, P. A., Kruse, A., & Schroder, J. (2012). Facial expression in {A}lzheimer's disease: Impact of cognitive deficits and neuropsychiatric symptoms. *American Journal of Alzheimer's Disease and Other Dementias, 27*(2), 100–106.

Simons, G., Pasqualini, M. C., Reddy, V., & Wood, J. (2004). Emotional and nonemotional facial expressions in people with Parkinson's disease. *Journal of the International Neuropsychological Society, 10*(4), 521–535.

Stel, M., van den Heuvel, C., & Smeets, R. C. (2008). Facial feedback mechanisms in autistic spectrum disorders. *Journal of Autism and Devevelopmental Disorders, 38*(7), 1250–1258.

Tron, T., Peled, A., Grinsphoon, A., & Weinshall, D. (2016). Automated Facial Expressions Analysis in Schizophrenia: A Continuous Dynamic Approach. In S. Serino, A. Matic, D. Giakoumis, G. Lopez, & P. Cipresso (Eds.). *Pervasive Computing Paradigms for Mental Health: 5th International Conference* (pp. 72–81). Cham: Springer International Publishing Revised Selected Papers.

Volker, M. A., Lopata, C., Smith, D. A., & Thomeer, M. L. (2009). Facial encoding of children with high-functioning autism spectrum disorders. *Focus on Autism and Other Development Disabilities, 24*(4), 195–204.

Yirmiya, N., Kasari, C., Sigman, M., & Mundy, P. (1989). Facial expressions of affect in autistic, mentally retarded and normal children. *Journal of Child Psychology and Psychiatry, 30*(5), 725–735.

Zhao, X., Dellandrea, E., Zou, J., & Chen, L. (2013). A unified probabilistic framework for automatic 3D facial expression analysis based on a Bayesian belief interference and statistical feature models. *Image and Vision Computing, 31*(3), 231–245.