This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering

A Feasibility Study of Autism Behavioral Markers in Spontaneous Facial, Visual, and Hand Movement Response Data

Manar D. Samad, Norou Diawara, Jonna L. Bobzien, John W. Harrington, Megan A. Witherow, and Khan M. Iftekharuddin *Senior Member, IEEE*

***Abstract*—Autism Spectrum Disorder (ASD) is a neurodevelopmental disability with atypical traits in behavioral and physiological responses. These atypical traits in individuals with ASD may be too subtle and subjective to measure visually using tedious methods of scoring. Alternatively, the use of intrusive sensors in the measurement of psychophysical responses in individuals with ASD may likely cause inhibition and bias. This work proposes a novel experimental protocol for non-intrusive sensing and analysis of facial expression, visual scanning, and eye-hand coordination to investigate behavioral markers for ASD. An institutional review board (IRB) approved pilot study is conducted to collect the response data from two groups of subjects (ASD and Control) while they engage in tasks of visualization, recognition, and manipulation. For the first time in the ASD literature, the facial action coding system (FACS) is used to classify spontaneous facial responses. Statistical analyses reveal significantly ($p<0.01$) higher prevalence of smile expression for the group with ASD with the eye-gaze significantly averted ($p<0.05$) from viewing the face in the visual stimuli. This uncontrolled manifestation of smile without proper visual engagement suggests an impairment in reciprocal social communication, e.g., social smile. The group with ASD also reveals poor correlation in eye-gaze and hand movement data suggesting deficits in motor coordination while performing a dynamic manipulation task. The simultaneous sensing and analysis of multimodal response data may provide useful quantitative insights into ASD to facilitate early detection of symptoms for effective intervention planning.**

Keywords*—Autism Spectrum Disorders, Psychophysical study, Behavioral marker, Facial action coding, Multimodal physiological measurement, Eye-tracking, Eye-hand coordination.

I. Introduction

AUTISM Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects multiple areas of an individual's verbal and nonverbal communication skills to varying extents. The onset of ASD occurs during the early developmental period of the child with the presentation of subtle deviances from typical behavioral and cognitive traits. Currently, there are no diagnostic biomarkers for ASD [1]; therefore, this disorder is commonly identified through direct visual observation of atypical behaviors.

However, visual observation methods have limitations in identifying subtle behavioral traits which may gradually lead to more complex behavioral impairments over the developmental period of the child with ASD. Therefore, a child diagnosed with ASD needs follow-up screenings to identify specific abnormalities in behavior and social communication skills. Clinicians and researchers have developed measurement tools involving rating scales, observational methods, and interviews to evaluate atypical traits among pre-school and school-aged children with ASD [2]. For example, Dawson et al. propose a measurement tool known as the Broader Phenotype Autism Symptom Scale (BPASS) [3] that requires trained clinicians to interact with children with ASD to observe and rate ASD-related traits on four items of expressiveness: eye contact, vocal prosody, facial expressions, and social smile. Although effective, the BPASS method entails significant time, cost, training, and expertise due to human interventions. An automated computer vision-based tool may assist in analyzing facial expressions and visual engagement to reduce time and expenses currently needed for screening behavioral markers in individuals with ASD. The behavioral markers gleaned from the physiology of the face and vision may further facilitate the computation of the severity and prognosis of the disorder.

This paper addresses the need for novel experimental methods to acquire multimodal response data that differentiate a group with ASD from a control group. Spontaneous physiological data about the facial expression, visual processing, and hand movement are simultaneously captured while the subjects engage in tasks involving visual stimuli. The simultaneous study of these modalities is expected to reveal more complex and correlated behavioral markers for the group with ASD, which may not be evident using unimodal data.

The remaining manuscript is organized as follows: Section 2 provides a background review on the existing experimental methods. Section 3 outlines the proposed experimental design, data and image processing frameworks to analyze multimodal physiological response data. Section 4 presents the outcomes of the proposed image processing framework and statistical inferences. A discussion following the results is provided in Section 5 to suggest potential behavioral markers for individuals with ASD. Section 6 concludes with future work.

M. D. Samad, M. A. Witherow and K. M. Iftekharuddin are with the Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, 23529 USA (e-mail: {msama005, mwith010, kiftekha}@odu.edu)

N. Diawara is with the Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529 USA (e-mail: ndiawara@odu.edu)

J. L. Bobzien is with the Department of Communication Disorders and Special Education, Old Dominion University, Norfolk, VA 23529 USA (e-mail: jbobzien@odu.edu)

J. W. Harrington is with the Department of Pediatrics, Eastern Virginia Medical School, Norfolk, VA 23507 USA (e-mail: john.harrington@chkd.org)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering

II. Background review

Activations of facial muscle, eye-contacts, and hand movement yield a variety of nonverbal expressions intended for inter-personal communication and social engagement. Nonverbal expressions convey about 80% of the cues in social communications [4]. Hence, impairments in social communication skills may manifest in multiple areas of nonverbal expressions used to regulate social interactions [5]. Social interactions entail reciprocal manifestation of communicative intents and expressiveness in the face by engaging the gaze toward an individual's face or reaching the hand toward the point of visual interest. Therefore, the simultaneous study of spontaneous responses from the face, eye-gaze, and hand motion may identify more intricate and subtle behavioral markers that are otherwise not obvious through visual screening. Prior studies have found eye-gaze metrics, such as gaze-fixation and fixation duration, useful in understanding the visual scanning mechanism of individuals with ASD [6], [7], [8]. The eye-gaze data may also provide with complementary information about spontaneous facial responses and coordination skills in hand movement while encountering visual stimuli. The existing eye-hand coordination studies are limited to static paradigms involving mouse-button press at the visual target [9] and do not capture the dynamic relationship between eye-gaze and hand motion. Several limitations of the existing studies are discussed in the following sections.

A. Facial expression analysis in ASD

The ASD literature generally investigates how individuals with ASD either recognize or respond to facial expressions displayed as visual stimuli. Facial responses in individuals with ASD can be attenuated and covert [10] or atypical [11], [12] when compared to those in healthy controls. Video frames of these facial responses are commonly evaluated by untrained human coders [13], [12], [11], which is prone to errors due to qualitative evaluations. The study of facial expression imitation skill in subjects with ASD [12] provides limited insights into their spontaneous mimicry of expressions or empathy toward the viewed emotion.

Spontaneous facial expressions in individuals with ASD are measured using electromyography (EMG) sensors. EMG sensors can capture subtle facial muscle movements that are otherwise imperceptible to human eyes. However, EMG probes are limited to a pair of facial muscles (typically cheek and eye-brow) [14], [15], [16] that may not distinguish responses from a variety of other facial actions. Furthermore, the steps for placing EMG electrodes, such as washing of facial skin with soap and alcohol swabs, may cause distress in subjects with ASD. The placement of electrophysiological sensors on different body parts such as face, fingers, shoulder, and chest of individuals [15] can significantly constrain the natural body, hand, and head movements of the subjects. Because over half of those diagnosed with ASD show anxiety and phobia related to novel experiences [17], these intrusive procedures and restrictions may overwhelm the subjects who are primarily children and eventually inhibit or bias their natural response data. Therefore, an ergonomic experimental setup is necessary for a psychophysical study to ensure minimal invasion and stress while maximizing engagement and elicitation of subjects with ASD. The concern regarding intrusiveness has motivated Hashemi et al. to replace wearable and intrusive devices with non-intrusive video cameras [18]. Consequently, this work proposes a hardware-software setup with non-intrusive imaging sensors to minimize physical constraints and interference with the subject. The facial imaging with computer vision methods are proposed to classify facial expressions into multiple action units using facial action coding system (FACS).

B. Facial action coding system

FACS provides a dictionary of unique facial muscle actions that are encoded in facial action units (FAUs) [19] to annotate different emotional and non-emotional states of the face. For example, upper lip raising (FAU 10) is found in disgusted facial expression [20] and the lip corner puller (FAU 12) represents a smile [21]. The facial dimple (FAU 14), contributed by the buccinator muscle, is reported as a marker of depression [22]. Lip pressing, annotated by FAU 24, represents resentment and lips apart (FAU 25) appears in surprised expression. One of the EMG electrodes is used to probe the activation of $Zygomaticus\ Major$ muscle [15], [16] in ASD, which is annotated by FAU 12. The manifestation of 'open mouth' [23] and the presence of depression and anxiety in individuals with ASD [17] can be annotated by FAU 25 and FAU 14 [22], respectively. Previously, computer vision-based detection of FAUs is used to identify anomalies in imitating facial expressions in subjects with schizophrenia [24]. This paper moves forward to study spontaneous FAUs in individuals with ASD instead of studying facial expression imitation skills.

C. Contributions

This work, for the first time in the literature, introduces computer vision methods to classify spontaneous facial action units in subjects with ASD. This method alleviates the shortcomings of subjective evaluations and intrusive sensing procedures in the literature. In this regard, a novel psychophysical experimental paradigm is proposed to engage the subjects, elicit their spontaneous responses, and simultaneously capture the response data using multimodal imaging sensors in a non-intrusive manner. The prevalence of FAUs is studied in conjunction with eye-tracker data to gauge the reciprocity in behavioral interaction. A correlation metric is proposed to evaluate the coordination of the eye-gaze fixation and mouse cursor movement during a dynamic manipulation task. The correlations between multimodal communication channels such as spontaneous facial expression, visual scanning, and hand motion are proposed to understand complex behavioral markers in subjects with ASD.

III. Methods

This section discusses the experimental design for human study and the image analysis framework

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering

(a) (b)

Fig. 1. Hardware-software platform for the study. (a) A subject sitting before the 68" TV display and camera sensors, (b) two monitors behind the scene to launch and coordinate the study.

A. Experimental design and protocol

The experimental setup for the proposed psychophysical study is carefully designed following several relevant studies [14], [25], [26], and in consultation and collaboration with physicians experienced with ASD. Institutional Review Boards from both Old Dominion University (ODU) and Eastern Virginia Medical School (EVMS) have approved the study protocol involving human subjects with and without ASD.

1) Subjects: Human subjects are recruited as follows: (i) a group exclusively diagnosed with high-functioning ASD with IQ score $>$70 and (ii) a healthy control group without a diagnosis of ASD. The age range of the participants is between seven and 20 years, which is suitable to undertake the tasks for the study. Among sixteen enrolled subjects, eight are diagnosed with ASD by professionals from the psychiatry practice at EVMS. Subjects with ASD have an average age of 13$\pm$4.4 years. The remaining eight subjects are healthy controls with an average age of 16$\pm$4.1 years. The two groups' ages have no significant difference (F(1, 14) = 2.66 with a large p-value (p$>$0.05)). Prior to enrollment, a potential subject is screened for inclusion and exclusion criteria via telephone. An eligible subject and parent review the informed consent and assent forms that outline the goals and tasks for the study. For the control group, the same inclusion and exclusion criteria are followed except the diagnosis of ASD or any disorder is considered an exclusionary criterion.

2) Hardware-software design: A hardware-software platform is developed to facilitate both ergonomic accommodation of the subjects and automatic operation of the sensors and stimuli. The platform includes a 68" multimedia television (TV) to display the visual stimulus as shown in Figure 1 (a). Lerner et al. place their subjects 2.5 feet from a 24" monitor [25], which we adjust to seven feet for the 68" TV display. Two monitors behind the scene (Figure 1 (b)) are used to launch and coordinate each trial in the study.

The two non-intrusive sensors used in this study are a video camera (Sony EVI-D70) and a *Mirametrix* S2 eye-tracker (www.mirametrix.com). Both sensors are positioned facing the subject's face in such a way that they do not obstruct the line of sight of the subject viewing the TV screen. The off-the-shelf eye-tracker system comes with a robust five-point calibration for the entire screen. According to the operating manual, the eye-tracker does not require a chin rest and allows freedom of head movement. During the calibration, the subject visually follows five bulls-eye shaped targets in a sequence as the application maps eye movements with corresponding positions on the screen. The calibration performance is displayed after each attempt and is repeated as many times as necessary until the eye-tracker system displays successful calibration. Successful calibration is confirmed for each subject prior to engaging in the actual study task. In addition, a graphical user interface (GUI) application is developed using Qt/C++ to collect the subject's perception of displayed facial expressions in each trial. The mouse cursor is programmed to automatically and sequentially initiate each step including the onset and end of a trial as shown in Figure 2.

TABLE I. A SUMMARY OF THE EXPERIMENTAL TASKS AND COLLECTED DATA.

Session	Duration per trial	Num. of Trials	Tasks	Collected Data
1	15 s	6 Expressions X 2 = 12 3D-faces	Visualization, recognition	Facial expression images, Eye-gaze
2	25 s	6 Expressions X 2 = 12 3D-faces	Visualization, manipulation	Facial expression Eye-gaze, Hand motion

3) Task and procedures: The subject is seated on an adjustable chair in front of the 68" TV with the computer mouse on a side table (Figure 1 (a)). Computer-based tasks involving and interacting with facial images have been found to be engaging and effective for studying individuals with ASD [27]. We use 3D faces with expressions as stimuli that offer greater sense of realism than 2D faces. Two tasks are designed for this study to induce better engagement and elicitation of spontaneous responses in subjects with ASD as follows.

The first task for the subjects is to recognize facial expression from 3D faces displayed as visual stimuli. Impairments in processing facial expressions are common symptoms for ASD [28], [14]. In subjects with ASD, emotional expressions can create unusual activation of certain regions of the brain [28], which may alter natural mimicry or empathy toward the displayed emotions. High resolution 3D faces, benchmarked with six prototypical expressions (happy, anger, fear, sad, surprise, and disgust) [29], are selected from the Binghamton 3D facial expression dataset to create 12 random trials using 12 different 3D faces with expressions. The sensors are triggered 1000 ms following the display of the visual stimulus since it takes 500-1000 ms to produce a facial response to a visual stimulus [14]. The subjects use the GUI to select their perception about the displayed 3D facial expression within a 15-second time constraint.

The second task entails manipulation (rotation, translation, scaling) of the visual stimuli (3D faces with expressions) using the mouse cursor (Figure 3 (b)). The orientation of the displayed face is known to affect the visual processing and perception of individual with ASD [30]. Accordingly, the subjects are instructed to adjust the orientation of the 3D face based on their preference while viewing the facial expression. This task is expected to yield spontaneous physiological response data in facial expressions, eye-gaze, and hand motion. The manipulation task involves a total of 12 random trials displaying 12 random 3D faces with expressions. A summary

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering

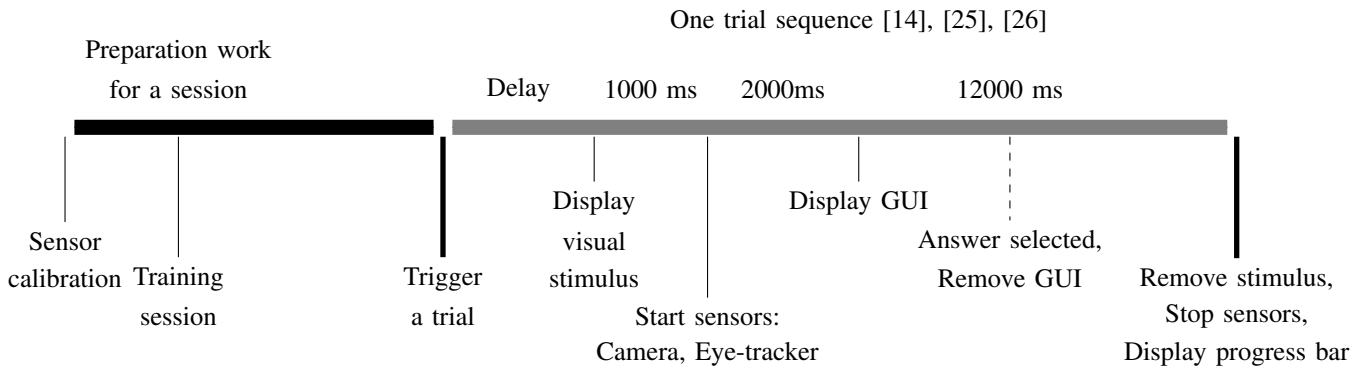
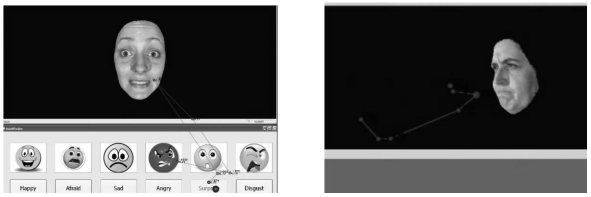


Fig. 2. Preparation work for a session and timing diagram for automated sequence of actions in a trial of the study.



(a) (b)

Fig. 3. Visual stimuli and experimental tasks. (a) Facial expression recognition task from 3D facial expression stimulus with the GUI at the bottom; (b) the manipulation of a 3D face with expression using the mouse cursor. The connected dots show the gaze paths while performing the task.

of the two tasks is presented in Table I. Subjects also perform several practice trials prior to starting the main session of each task. Break periods are allowed in between the two tasks or at any time they request a break.

4) Data collection: The proposed hardware-software system automatically starts and stops the video camera and eye-tracker to simultaneously capture facial image frames and eye-tracker data respectively over the entire period of a trial. The eye-tracker provides videos of the eye-gaze map overlaid on the visual scene, time-sampled coordinate locations and duration of individual gaze fixations, and the mouse cursor coordinate position mapped on the same TV display.

B. Analysis of facial responses

A publicly available dataset of FAU, known as the Bosphorous facial expression dataset [31], is used to train FAU classifier models. Prototypical facial expressions can consist of more than one FAU and the likelihood of neutral facial expression may be high considering the subtlety in spontaneous facial expressions. Therefore, we contrast five FAUs (introduced in Section II-B) from the neutral expression using five binary classifiers: FAU 10 vs. neutral, FAU 12 vs. neutral, FAU 14 vs. neutral, FAU 24 vs. neutral, and FAU 25 vs. neutral. The five binary classifier models are trained to separately probe the prevalence of five FAUs from the same input facial image. The image processing pipeline for extracting facial image features is discussed in the following sections.

Preprocessing: A total of 24 video clips are obtained from 24 trials following the two tasks. Each video clip is sampled at the rate of one frame per second to yield 15 facial image frames from a 15-second video clip. A robust facial landmark detection algorithm [32] is used to detect facial landmarks (eye, nose, lips, facial edges) that are subsequently used to extract the facial region from a video frame and to perform a rigid registration to spatially align facial images with similar regions of a reference facial image. The registration step establishes the spatial correspondence in facial features for the subsequent segmentation and feature extraction steps. The registered facial images are then resampled and normalized into N-by-N pixel dimension to obtain a common topology for feature extraction.

Facial Segmentation: Following the registration, we segment the lower half of the face to account for the features of our target FAUs similar to [33]. Note that our five target FAUs appear only in the lower half of the face which is more sensitive to subtle facial expressions than its upper half [34].

Feature extraction: The segmented lower region of the facial image is converted to gray scale from color. A bank of 2D Gabor functions are used to extract features corresponding to five scales and eight orientations. A total of 40 Gabor filters have been constructed to extract Gabor features from the segmented lower face. The state-of-the-art studies on FAU recognition have used similar Gabor features [35], [33].

Feature selection: A feature selection step obtains a smaller subset of the extracted Gabor features. First, ground truth facial images from the Bosphorous dataset are used to identify a subset of Gabor features that yield highest mean squared differences between a FAU and the neutral expression. The mean squared differences (MSD) of Gabor features for a particular FAU are obtained over the images samples (N = 50) available in the Bosphorous dataset as follows.

$$MSD = \frac{1}{N}\sum_{k=1}^{N}\left[F_N(k,j) - F_{FAU}(k,j)\right]^2, \qquad (1)$$

where, F_N and F_{FAU} represent Gabor feature vectors for neutral expression and a particular FAU, respectively and k is the index for the image samples in the ground truth dataset. Gabor features are ranked based on their MDS values and the feature indices with MSD values higher than a threshold

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering

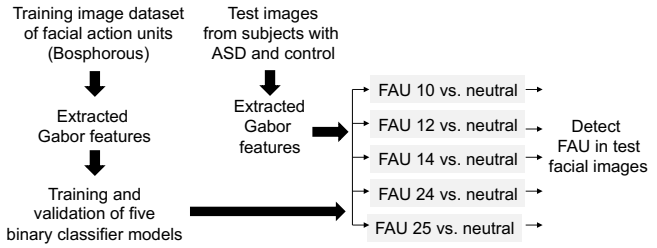


Fig. 4. Steps for classification and evaluation of the probe facial images through five binary classifier models corresponding to five target FAUs, respectively.

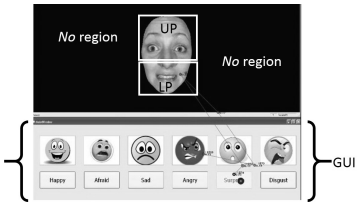


Fig. 5. Four visual areas of interest (AOI) for eye-tracking: (a) UP represents the upper part of face; (b) LP represents and lower part of the face; and (c) GUI represents the graphical user interface region. The remaining regions belong to the No region.

are stored for the corresponding binary classifier. The stored indices for five binary classifiers are later used to select features from test facial images prior to the classification.

1) Evaluation: The selected Gabor features from Section III-B are then used for training and testing the binary classifiers as shown in Figure 4 .

First, each binary classifier is trained and validated using the selected Gabor features obtained from the Bosphorous dataset and a ten-fold cross-validation scheme. Based on our prior work [36], we use support vector machine (SVM) with radial basis function (RBF) kernel as the classifier. The cross-validation step is also used to optimize the classifier model parameters to yield the best classification performance. Second, the performance and optimization results from the cross-validation step are considered to construct five binary classifier models as introduced in Section III-B and shown in Figure 4. Third, the trained binary classifier models are used to detect the prevalence of five FAUs in the test facial images obtained from the two groups of subjects. For a subject, the percentage of prevalence of a particular FAU is measured as follows,

$$PFAU\_X = \frac{NAU}{TNFI} \times 100, \tag{2}$$

where, NAU is the number of facial images detected with FAU = 'X', TNFI stands for total number of available facial images of the test subject, and X denotes one of the five target FAUs: 10, 12, 14, 24, and 25. The PFAU_X represents the probability of a certain FAU to appear in a subject's face while participating in the study. Between-group and within-group designs of ANOVA tests are performed to find if any of the five FAUs shows significant prevalence as a differential marker for the group with ASD. We consider a significance level α=0.05 for the ANOVA tests.

C. Analysis of eye-tracker data

The eye-tracker system provides time-sampled coordinate locations for both gaze fixations and the mouse cursor. A gaze fixation with equal or more than 320 ms duration is termed as voluntary gaze fixation [37], which are considered in this study to ascertain the point of visual focus and to avoid any inadvertent gaze fixations. The voluntary gaze fixations are grouped into four visual areas of interest (AOI) in the visual scene as shown in Fig. 5. The percentage of voluntary gaze fixation duration (PVGFD) at each of four areas is defined as follows,

$$PVGFD = \frac{VGFDA}{TVGFD} \times 100. \tag{3}$$

Here, VGFDA is the voluntary gaze fixation at an AOI and TVGFD is total voluntary gaze fixations during the trial. For each subject, this percentage is averaged over the number of trials in the study. Within-group and between-group designs of ANOVA are used with a significance level α=0.05. The eye-tracker also provides time-sampled locations of the mouse cursor with the gaze fixations mapped on the same visual coordinate. Hence, the mouse cursor movement can be correlated with eye-gaze fixation data to yield useful metric about the eye-hand coordination during the manipulation task.

D. Experimental design and sample size

Each of the six facial expression stimuli can be assumed independent. While there is repetition in the expressions due to random selection of the trials, the six facial expressions in the two tasks yield 12 independent trials. Therefore, the observations from these independent trials within the same subject are assumed uncorrelated. Hence, the total sample size of response data is 192 from 16 subjects (divided into two groups) in response to 12 independent trials. Statistical power analysis ensures that the sample size is large enough to validate the results with a power of more than 95%.

IV. Experimental results

The results following the proposed steps in Section III are discussed below.

A. Processing of facial image

The subject's head is not constrained to allow spontaneous participation in the study. Consequently, there are facial image frames with posed faces, facial occlusions by the hand, and partial faces. The facial images with out-of-plane head rotation or occlusions are carefully discarded to eliminate error in the classification. The faces with in-plane rotations are adjusted via a landmark-based registration step as discussed in Section III-B. Figures 6 (a) and (b) show the landmarks on the reference facial image and the test image, respectively. Figure 6

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering

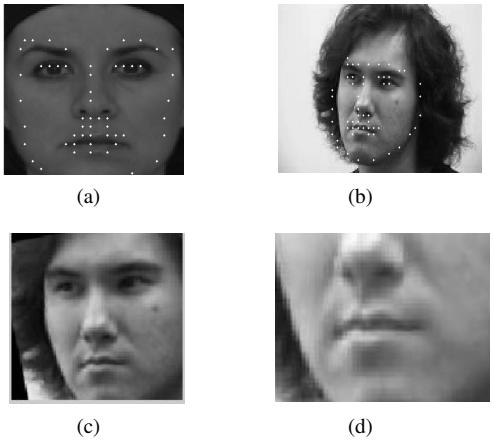


Fig. 6. (a) Ground truth reference face with facial landmarks, (b) test face with detected landmarks, (c) registered face after rigid transformation of landmarks, (d) segmented lower region of the face.

TABLE II. ANOVA TEST RESULTS ON THE PERCENTAGE OF PREVALENCE OF FAUS FROM FACIAL IMAGES.

	Within-group design	
	Action Units	Subject
Within Control	F(4, 35) = 10.5 (p<0.001)	F(7, 32) = 0.53 (p>0.05)
Within ASD	F(4, 35) = 6.45 (p<0.001)	F(7, 32) = 1.21 (p>0.05)
Action Units	Between-group design	
AU 10	F(1, 14) = 2.99	p>0.05
AU 12	F(1, 14) = 9.23	p<0.01
AU 14	F(1, 14) = 0.95	p>0.05
AU 24	F(1, 14) = 0.45	p>0.05
AU 25	F(1, 14) = 0.81	p>0.05

(c) shows the test image after the rigid registration. Note that the mouth region of the registered face in Figure 6(c) is at the similar pixel locations of that of the reference face in Figure 6 (a). Figure 6 (d) shows the segmented lower part of the test face after the conversion to gray scale. The dimensions of the registered image and the segmented lower part of the image are 60x60 and 40x60, respectively. The Gabor feature extraction step yields a total of 96,000 feature points from the lower part of the facial image. A small subset of representative features are selected for classification for training each binary classifier using Eq. 1.

B. Cross-validation of training data

A ten-fold cross validation evaluates the performance of the five classifier models using the Bosphorous dataset. The areas under the receiver operating characteristics curve are found to be 0.95, 0.97, 0.94, 0.88, and 0.96 for classifying FAU 10, 12, 14, 24, and 25 against the neutral expression, respectively.

C. Prevalence of FAUs

The video camera yields more than 100 facial image frames per subject on average and a number of image frames are discarded because of ill-pose and occlusions. The remaining facial images are tested using the five trained binary classifier models. The average percentage of prevalence (APP) of FAUs for individual subjects is considered for further ANOVA tests

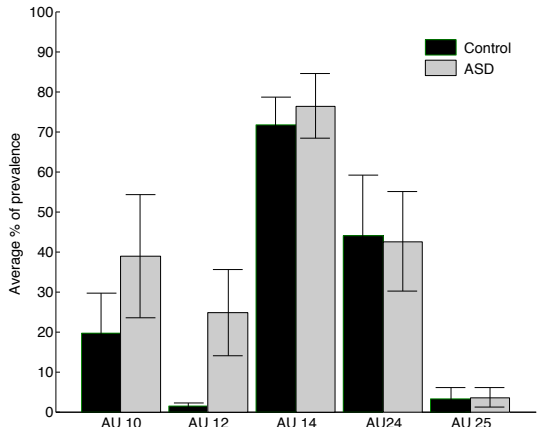


Fig. 7. Average percentage of prevalence (APP) of the five FAUs for the control group and the group with ASD.

TABLE III. ANOVA TEST RESULTS. THE UPPER TRIANGLE IS FOR THE GROUP WITH ASD. THE LOWER TRIANGLE IS FOR THE CONTROL GROUP. THE SIGNIFICANT INFERENCES ARE HIGHLIGHTED. NO= No REGION, UP= UPPER PART, LP= LOWER PART OF THE FACE, GUI=GRAPHICAL USER INTERFACE, PVGD=PERCENTAGE OF VOLUNTARY GAZE DURATION.

	Within-group design			
	No	UP	LP	GUI
No	X	F=19.95, **p<0.001**	F=13.72 **p<0.001**	F=1.4, p>0.05
UP	F=2.12 p>0.05	X	F=0.54 p>0.05	F=74.21 **p<0.001**
LP	F=3.63 p>0.05	F=6.64 **p<0.05**	X	F=48.21, **p<0.001**
GUI	F=23.26 **p<0.001**	F=1.48 p>0.05	F=43.22, **p<0.001**	X
	Between-group design			
PVGD	No	UP	LP	GUI
F=4.06, p=0.063	F=6.39, **p<0.05**	F=7.54, **p<0.05**	F=0.35, p>0.05	F=1.3, p>0.05

and the results (Table II). A within-group ANOVA test shows significant difference among APPs of the five FAUs for the control group, F(4, 35) =10.5, (p<0.001) as well as for the group with ASD, F(4, 35) = 6.45, (p<0.001). Figure 7 shows a comparative illustration of the two groups in terms of APP of five FAUs. Table II shows that the group with ASD exhibits significantly higher prevalence of FAU 12, F (1, 14) = 9.23, (p<0.01) than the control group.

D. Gaze fixation and facial response

The percentage of time spent on voluntary gaze fixation or the percentage voluntary gaze duration (PVGD) (Figure 8(a) and Table III), has been found to be higher within the control group (74%) when compared to the group with ASD (68%) with no statistical difference between the two groups (F(1, 14)= 4.06, p>0.05). A lower PVGD in subjects with ASD may indicate their inefficiency in visual scanning.

The upper triangle of Table III shows significantly higher percentage of gaze fixation duration (GFD) in scanning the GUI and the No region than the facial regions (lower and upper face) for the group with ASD. A significantly higher percentage of GFD is found at the No region when compared

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering

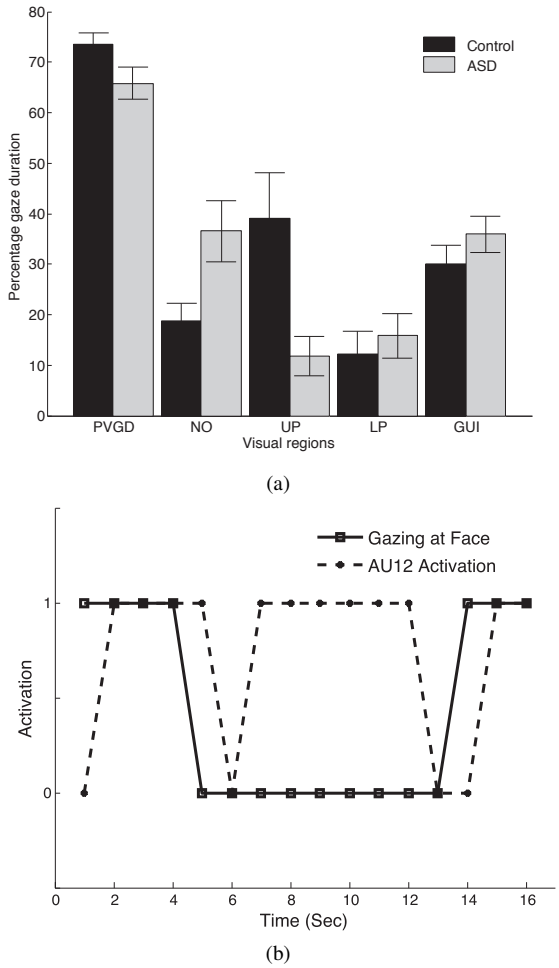


Fig. 8. (a) Percentage of gaze duration at four visual scenes (No = No region, UP= upper face, LP= lower face, GUI= Graphical User Interface). PVGD= Percentage of Voluntary Gaze Duration. (b) Activation of FAU 12 vs time and fixating gaze on the face vs time a subject with ASD view a face with happy expression. '1' indicates fixating gaze on the face and '0' indicates gazing elsewhere.

to the upper part (F=19.95, $p<0.001$) and lower part (F=13.72, $p<0.001$) of the face. It is evident that individuals with ASD avoid visual interaction with the human face and exert more visual scanning elsewhere. On the other hand, the control group shows opposite results as shown in the lower triangle of Table III. The control group spends a significantly higher percentage of GFD in scanning the upper part of the face than the lower part (F=6.64, $p<0.05$) and greater percentage of time processing the GUI than scanning the No region (F=23.26, $p<0.001$) or the lower part of the face (F=43.22, $p<0.001$). A between-group ANOVA test indicates that the control group spends higher percentage of time visually scanning the upper part of the face when compared to the group with ASD (F=7.54, $p<0.05$). The group with ASD spends a significantly higher percentage of time scanning the No region when compared to the control group (F=6.39, $p<0.05$).

The eye-gaze fixation may further help to explain the reciprocity in facial response such as social smile. The lack of visual engagement with the face along with uncontrolled manifestation of FAU 12 or smile expressions (Figure 7) may be indicative of impairments in reciprocal communication skills for the group with ASD. Figures 8 (c) and (d) illustrate temporal activation data for FAU 12 and gaze fixation for two representative subjects with ASD while the visual stimulus is a face with happy expression. The video frames with the subject's face and eye-gaze fixations are sampled at one-frame per second to show negative correlation between the activation of FAU 12 and gaze fixation on the face.

E. Eye-Hand coordination

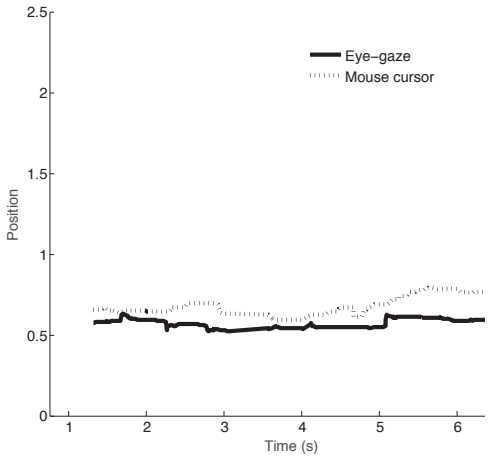
The manipulation task entails frequent movement of the mouse cursor while viewing and manipulating the 3D face with expression. The eye-tracker system yields spatial and temporal data related to the eye-gaze fixations and hand movement via the mouse cursor position. The correlation between the eye-gaze fixation and mouse cursor position over time can offer important insight into the motor and coordination skill of subjects with ASD. The Euclidean distance between the 2D (X-Y) coordinate position and the origin is used to represent the relative position for eye-gaze fixations or the mouse cursor. Figure 9 shows representative illustrations for the eye-gaze and mouse cursor positions over time for the two groups. The subject with ASD yields frequent movement in the mouse cursor position that suggests an incoherent hand movement with respect to the eye-gaze fixation. This incoherence between the hand movement and eye-gaze may affect the dexterity or eye-hand coordination skill. The correlation coefficient between time-sampled activations of eye-gaze and hand movements is lower in the group with ASD when compared to the control group in response to positive (0.0739 ± 0.350 vs. 0.234 ± 0.365) and negative facial expressions (0.194 ± 0.398 vs. 0.259 ± 0.353).

V. Discussion

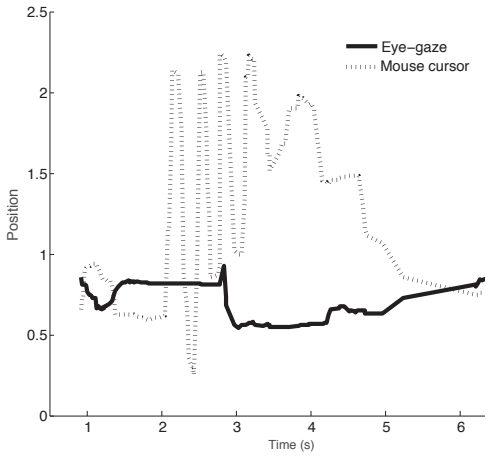
This work demonstrates systematic acquisition and quantitative evaluations of multi-modal physiological responses in the investigation of behavioral biomarkers for subjects with ASD. The multi-modal physiological data offer useful complementary information in the interpretation of complex nonverbal expressions of the group with ASD. Our major findings for the group with ASD include: 1) individuals with ASD may have uncontrolled manifestation of FAU 12, which is the major constituent of smile expression; 2) spontaneous facial responses are not synchronized with their visual engagement with facial expressions; 3) a poor correlation in dynamic eye-hand movements is found during the manipulation task.

The lack of perception of negative emotional expressions, such as anger, fear, may contribute to the manifestation of incongruent smile expression while viewing negative expressions [38]. Hence, FAU 12 or smile expression may be a useful target for identifying ASD-related behavioral traits. The smile expression is easier to elicit and smiling is known to be one of the common expressions in human infants, which is annotated by the baby FACS [39]. Additionally, human infants

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2768482, IEEE Transactions on Neural Systems and Rehabilitation Engineering



(a)



(b)

Fig. 9. Relative positions (Euclidean distance from the origin of the visual scene) for the Eye-gaze fixation and mouse cursor versus time for (a) the control group and (b) the group with ASD.

may manifest social smile as early as six months of age by making face-to-face interaction with adults [40]. Social smile is a reciprocal behavior that involves simultaneous eye-contact while a smile is shared between two individuals. Computer vision methods may be used to detect FAU 12 to subsequently identify anomalies in social smile during early developmental period of child to facilitate early screening and intervention planning. Furthermore, the detection of FAU 12 may be used to study the activation of the underlying *Zygomaticus Major* muscle without using intrusive EMG electrodes. This may improve prior studies [16], [14], which may have resulted in inconclusive findings due to intrusive methods of probing facial response.

The higher prevalence of FAUs in the group with ASD when compared to the control group (Figure 7) may contradict a general notion that subjects with ASD are unable to produce facial expressions. This study enrolls high-functioning subjects with ASD who are known to engage and respond more appropriately than their low-functioning peers when the stimulus or the task is arousing [41], [11]. Our previous study on non-intrusive 3D imaging of facial response also reveals higher facial muscle activation with the same groups of subjects [42].

The group with ASD spends a significantly more time gazing at visual scenes other than the displayed face with expression (Table III). This phenomenon has contributed to a negative correlation between the elicited FAU 12 and visual engagement (Figure 8 (b)). In contrast to the control group, the lack of synchronism between facial expression and visual engagement suggests a behavioral marker for ASD, which may be automatically screened by simultaneously imaging facial expression and eye-gaze fixation. Individuals with ASD further show that the mouse cursor driven movement of the 3D face goes beyond the region of their visual focus to possibly avoid gaze-fixations on the 3D face. The averted eye-gaze may eventually contribute to lower eye-hand correlation, which, in turn, may affect the motor coordination and dexterity of an individual.

In contrast to the group with ASD, the low prevalence of FAU 12 in control subjects may be due to a lack of enthusiasm in trivial tasks with facial expressions. The eye-tracker reveals that control subjects are visually well focused on the upper part of the face (eye-contact region) and the GUI while attempting to recognize the facial expression. Hence, the control group may have more control of their eye-gaze and emotional expressions than the group with ASD. Individuals who are diagnosed with neurological disorders like ASD are known to have less control over their physiological expressions [43]. Emotional expressions such as smiling and crying can be induced involuntarily even in cases of palsy or lack of emotional control.

A quantitative and simultaneous evaluation of FAUs, eye-tracking data, and hand movement via computer vision-based processing, as proposed in this study, reveals several differential traits as potential behavioral markers for individuals with ASD. Our proposed computer-based environment and non-intrusive sensing are also found to be useful in engaging subjects with ASD in repetitive tasks. Therefore, computer-based virtual environments [44] may be used to stimulate the anomalies detected as behavioral markers so that an appropriate and early intervention can be arranged.

VI. Conclusions and Future Work

This paper presents a computer vision-based novel psychophysical study to investigate behavioral markers for individuals with ASD. The simultaneous study of spontaneous responses in the face, vision, and hand movement reveal several differential traits for the group with ASD. The uncontrolled manifestation of smile expression without visually engaging with facial expressions suggests an atypical behavior due to impairment in the communication skill of individuals with ASD. The low correlation between eye-gaze and hand movement suggests poor coordination skill for subjects with ASD. These behavioral markers may be objectively measured in children with ASD during their pre-school ages to facilitate early screening and intervention planning.

Although the results of this pilot study are promising, a full-scale study on a larger population is needed to understand the variability and dynamics of the observed behavioral marker. The future work aims to employ a social stimulus involving animated 3D face with expression and voice to capture temporal facial response data.

ACKNOWLEDGMENT

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

[1] W. Mandy, T. Charman, K. Puura, and D. Skuse, "Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: evidence from Finnish and UK samples," *Autism*, vol. 18, no. 1, pp. 45–54, Jan 2014.

[2] J. A. Deisinger, *The Broad Autism Phenotype*. Emerald Group Publishing Ltd., March 2015.

[3] G. Dawson, A. Estes, J. Munson, G. Schellenberg, R. Bernier, and R. Abbott, "Quantitative assessment of autism symptom-related traits in probands and parents: Broader Phenotype Autism Symptom Scale," *J Autism Dev Disord*, vol. 37, no. 3, pp. 523–536, Mar 2007.

[4] J. J. Thompson, *Beyond words: Nonverbal communication in the classroom.* New York: MacMillan Publishing Company, April 1973.

[5] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders.* Washington, DC: , 2000.

[6] J. B. Wagner, S. B. Hirsch, V. K. Vogel-Farley, E. Redcay, and C. A. Nelson, "Eye-tracking, autonomic, and electrophysiological correlates of emotional face processing in adolescents with autism spectrum disorder," *J Autism Dev Disord*, vol. 43, no. 1, pp. 188–199, Jan 2013.

[7] U. Lahiri, Z. Warren, and N. Sarkar, "Design of a gaze-sensitive virtual social interactive system for children with autism," *IEEE Trans Neural Syst Rehabil Eng*, vol. 19, no. 4, pp. 443–452, Aug 2011.

[8] T. Dratsch, C. Schwartz, K. Yanev, L. Schilbach, K. Vogeley, and G. Bente, "Getting a grip on social gaze: control over others' gaze helps gaze detection in high-functioning autism," *J Autism Dev Disord*, vol. 43, no. 2, pp. 286–300, Feb 2013.

[9] A. Crippa, S. Forti, P. Perego, and M. Molteni, "Eye-hand coordination in children with high functioning autism and Asperger's disorder using a gap-overlap paradigm," *J Autism Dev Disord*, vol. 43, no. 4, pp. 841–850, Apr 2013.

[10] D. Fein, *The neuropsychology of Autism.* Oxford University Press, April 2011.

[11] D. J. Faso, N. J. Sasson, and A. E. Pinkham, "Evaluating posed and evoked facial expressions of emotion from adults with autism spectrum disorder," *J Autism Dev Disord*, vol. 45, no. 1, pp. 75–89, Jan 2015.

[12] M. A. Volker, C. Lopata, D. A. Smith, and M. L. Thomeer, "Facial encoding of children with high-functioning autism spectrum disorders," *Focus on Autism and other development disabilities*, vol. 24, no. 4, pp. 195–204, Oct 2009.

[13] M. Stel, C. van den Heuvel, and R. C. Smeets, "Facial feedback mechanisms in autistic spectrum disorders," *J Autism Dev Disord*, vol. 38, no. 7, pp. 1250–1258, Aug 2008.

[14] P. M. Beall, E. J. Moody, D. N. McIntosh, S. L. Hepburn, and C. L. Reed, "Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder," *J Exp Child Psychol*, vol. 101, no. 3, pp. 206–223, Nov 2008.

[15] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder," *Int. J of Human-Computer Studies*, vol. 66, no. 9, pp. 662 – 677, 2008.

[16] A. Rozga, T. Z. King, R. W. Vuduc, and D. L. Robins, "Undifferentiated facial electromyography responses to dynamic, audio-visual emotion displays in individuals with autism spectrum disorders," *Dev Sci*, vol. 16, no. 4, pp. 499–514, Jul 2013.

[17] J. W. Harrington and K. Allen, "The clinician's guide to autism," *Pediatrics in review*, vol. 35, no. 2, pp. 62–78, Feb. 2014.

[18] J. Hashemi, M. Tepper, T. Vallin Spina, A. Esler, V. Morellas, N. Papanikolopoulos, H. Egger, G. Dawson, and G. Sapiro, "Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants," *Autism Res Treat*, vol. 2014, 2014.

[19] P. Ekman and W. Friesen, "Facial action coding system (FACS): A technique for the measurement of facial action, Palo Alto, CA: Consulting," 1978.

[20] T. Kanade, J. Cohn, and Y. L. Tian, "Comprehensive database for facial expression analysis," *IEEE Int. Conf. Autom. Face Gesture Recogn.*, p. 46, 2000.

[21] M. Hoque, D. McDuff, and R. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," *IEEE Trans. on Affective Computing*, vol. 3, no. 3, pp. 323–334, July 2012.

[22] J. Cohn, T. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. Padilla, F. Zhou, and F. De la Torre, "Detecting depression from facial actions and vocal prosody," in *3rd Int. Conf. on Affective Computing and Intelligent Interaction*, Sept 2009, pp. 1–7.

[23] H. Ozgen, G. S. Hellemann, R. K. Stellato, B. Lahuis, E. van Daalen, W. G. Staal, M. Rozendal, R. C. Hennekam, F. A. Beemer, and H. van Engeland, "Morphological features in children with autism spectrum disorders: a matched case-control study," *J Autism Dev Disord*, vol. 41, no. 1, pp. 23–31, Jan 2011.

[24] J. Hamm, C. G. Kohler, R. C. Gur, and R. Verma, "Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders," *J. Neurosci. Methods*, vol. 200, no. 2, pp. 237–256, Sep 2011.

[25] M. D. Lerner, J. C. McPartland, and J. P. Morris, "Multimodal emotion processing in autism spectrum disorders: an event-related potential study," *Dev Cogn Neurosci*, vol. 3, pp. 11–21, Jan 2013.

[26] K. Edwards, "The face of time: Temporal cues in facial expressions of emotion." *Psychological Science*, vol. 9, pp. 270–276, 1998.

[27] I. M. Hopkins, M. W. Gower, T. A. Perez, D. S. Smith, F. R. Amthor, F. C. Wimsatt, and F. J. Biasini, "Avatar assistant: improving social skills in students with an ASD through a computer-based intervention," *J Autism Dev Disord*, vol. 41, no. 11, pp. 1543–1555, Nov 2011.

[28] M. Dapretto, M. S. Davies, J. H. Pfeifer, A. A. Scott, M. Sigman, S. Y. Bookheimer, and M. Iacoboni, "Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders," *Nat. Neurosci.*, vol. 9, no. 1, pp. 28–30, Jan 2006.

[29] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," *7th Int. Conf. on Autom. Face and Gesture Recog.*, pp. 211–216, Apr. 2006.

[30] S. Weigelt, K. Koldewyn, and N. Kanwisher, "Face identity recognition in autism spectrum disorders: a review of behavioral studies," *Neurosci Biobehav Rev*, vol. 36, no. 3, pp. 1060–1084, Mar 2012.

[31] A. Savran and L. Akarun, "Bosphorous database for 3D face analysis," *Proc. first COST 2101 workshop biometrics and identity management*, pp. 47–56, May 2008.

[32] X. Zhu and D. Ramanan, "Face detection, pose estimation and landmark localization in the wild," *Computer Vision and Pattern Recognition (CVPR) Providence*, June 2012.

[33] A. Savran, B. Sankur, and M. T. Bilge, "Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units," *Pattern Recog.*, vol. 45, pp. 767–782, 2012.

[34] M. D. Samad and K. M. Iftekharuddin, "Radial curve-based classification and localization of features for 3D facial expressions," *Proc. SPIE Three-Dimensional Imaging, Visualization, and Display*, vol. 9117, June 2014.

[35] T. Wu, N. J. Butko, P. Ruvolo, J. Whitehill, M. S. Bartlett, and J. R. Movellan, "Multilayer Architectures for Facial Action Unit Recognition," *IEEE Trans Syst Man Cybern B Cybern*, vol. 42, no. 4, pp. 1027–1038, August 2012.

[36] M. D. Samad and K. M. Iftekharuddin, "Frenet frame-based generalized space curve representation for pose-invariant classification and recognition of 3-D face," *IEEE Trans. on Human-Machine Sys.*, vol. 46, no. 4, pp. 522–533, Aug 2016.

[37] W. Graf and H. Krueger, "Ergonomic evaluation of user-interfaces by means of eye-movement data," *Proc. of the 3rd Int. Conf. on Human-computer Interaction*, pp. 659–665, 1989.

[38] K. Humphreys, N. Minshew, G. L. Leonard, and M. Behrmann, "A fine-grained analysis of facial expression processing in high-functioning adults with autism," *Neuropsychologia*, vol. 45, no. 4, pp. 685–695, Mar 2007.

[39] H. Oster, "Emotion in the infant's face: insights from the study of infants with facial anomalies," *Ann. N. Y. Acad. Sci.*, vol. 1000, pp. 197–204, Dec 2003.

[40] R. N. Emde, T. J. Gaensbauer, and R. J. Harmon, "Emotional expression in infancy; a biobehavioral study," *Psychol. Issues*, vol. 10, no. 01, pp. 1–200, 1976.

[41] D. Mathersul, S. McDonald, and J. A. Rushby, "Automatic facial responses to affective stimuli in high-functioning adults with autism spectrum disorder," *Physiol. Behav.*, vol. 109, pp. 14–22, Jan 2013.

[42] M. D. Samad, J. L. Bobzien, J. W. Harrington, and K. M. Iftekharuddin, "Non-intrusive optical imaging of face to probe physiological traits in autism spectrum disorder," *Optics and Laser Technology* , vol. 77, pp. 221 – 228, 2016.

[43] B. Atchison and D. Dirette, *Conditions in Occupational Therapy: Effect on Occupational Performance*. Lippincott Williams and Wilkins, Nov. 2011.

[44] U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, and N. Sarkar, "Design of a virtual reality based adaptive response technology for children with autism," *IEEE Trans Neural Syst Rehabil Eng*, vol. 21, no. 1, pp. 55–64, Jan 2013.