ON QUANTIFYING FACIAL EXPRESSION-RELATED ATYPICALITY OF CHILDREN WITH AUTISM SPECTRUM DISORDER

Tanaya Guha[1], *Zhaojun Yang*[1], *Anil Ramakrishna*[1], *Ruth B. Grossman*[2,3]
Darren Hedley[2], *Sungbok Lee*[1], *Shrikanth S. Narayanan*[1]

[1]Signal Analysis and Interpretation Lab, University of Southern California, Los Angeles, CA
[2] Emerson College and [3]University of Massachusetts Medical School, Boston, MA

ABSTRACT

Children with Autism Spectrum Disorder (ASD) are known to have difficulty in producing and perceiving emotional facial expressions. Their expressions are often perceived as *atypical* by adult observers. This paper focuses on data driven ways to analyze and quantify atypicality in facial expressions of children with ASD. Our objective is to uncover those characteristics of facial gestures that induce the sense of perceived atypicality in observers. Using a carefully collected motion capture database, facial expressions of children with and without ASD are compared within six basic emotion categories employing methods from information theory, time-series modeling and statistical analysis. Our experiments show that children with ASD exhibit lower complexity in facial dynamics, with the eye regions contributing more than other facial regions towards the differences between children with and without ASD. Our study also notes that children with ASD exhibit lower left-right facial symmetry, and more uniform motion intensity across facial regions.

Index Terms— Affect, Autism, Emotion, Facial expressions, Motion Capture.

1. INTRODUCTION

Facial expressions provide non-verbal manifestations of internal emotional states that play a critical role in interpersonal communication and social interactions. Children with Autism Spectrum Disorder (ASD), who usually have restrictive social-communication skills, are known to have difficulty in producing and perceiving emotional facial expressions [1, 2]. Their expressions are often perceived as *atypical* or *awkward* as compared to their typically developing (TD) peers by typical adult observers. This perception of awkwardness is holistic, and a clinically acceptable qualitative measure of Autism [3]. Understanding the fine details of facial expression production mechanisms of children with ASD can bring objective insights into the nature of the perceived awkwardness.

Psychological work has established links between children with ASD and atypicality in their facial gestures, prosody, and body gestures [4, 5, 6, 7]. On the computational front, effort has been made to analyze atypicality in prosody [8, 9] and asynchronization of speech and body gestures of children with ASD [5, 10]. Computational work to analyze and quantify subtle differences in facial expressions that are otherwise difficult to understand by mere visual inspection is scarce, but nevertheless of great importance.

Motion capture (MoCap) data analysis was introduced as a powerful approach for quantifying differences in facial expressions between ASD and TD groups in our previous work [11]. In this a preliminary study [11], subjects with autism were found to have more rough facial and head motion.

In this paper, we investigate the emotion-specific atypicality in facial expressions of children with ASD using a larger MoCap database, by looking at global as well as region-based facial movements and dynamics. To this end, we group facial expressions into six basic emotion categories (*Anger, Disgust, Fear, Happiness, Sadness and Surprise*), and study how the characteristics of facial gestures vary with the emotions being conveyed. Our goal is two-fold: (i) understanding the overall complexity of the underlying mechanisms that generate facial expressions; (ii) examining the divergence between ASD and TD subjects in terms of region-based dynamics and activation of emotion-specific expressions. To achieve this, we employ various methods from information theory, statistics and time-series modeling. Characteristics of each emotion group are examined separately by analyzing facial MoCap marker data at two spatial scales using the entire face, and the eight local regions that divide a face (see Fig. 1).

2. THE MIMICRY DATABASE

This paper uses a MoCap marker database (designed and created by R. Grossman at the Facelab [12]) that has 45 subjects (24 with ASD and 21 TD) aged between 9 to 14 years. The subjects were shown emotional facial expression videos (reference stimuli) from the Mind Reading CD - a psychology resource [13]. The subjects were instructed to mimic those expressions. There are two predefined, very similar sets of expressions with 18 tasks in each set. Each subject mimics only one set of expressions, i.e. , 18 different expressions. These expressions include smiling, frowning, being tearful, etc., and belong to one of the six basic emotion groups - Anger, Fear, Disgust, Happiness, Sadness and Surprise.

Data were collected from 32 facial markers worn by each child (as shown in Fig. 1) using 6 MoCap cameras at 100 fps. Four stability markers were placed on the forehead and ears, and are used to measure and correct head motion. The positions of the remaining 28 markers are recomputed with respect to the stability markers to factor out movement caused by head motion so that we can focus on expression-related motion. Information from these 28 markers is used for analysis of facial expressions.

3. DATA ANALYSIS AND RESULTS

Facial MoCap data were subject to proper alignment, artifact removal, missing data interpolation, smoothing, and face normalization as detailed in [11]. Face normalization is important because it removes subject-specific variability due to differences in facial

This work was funded by NIH and NSF.

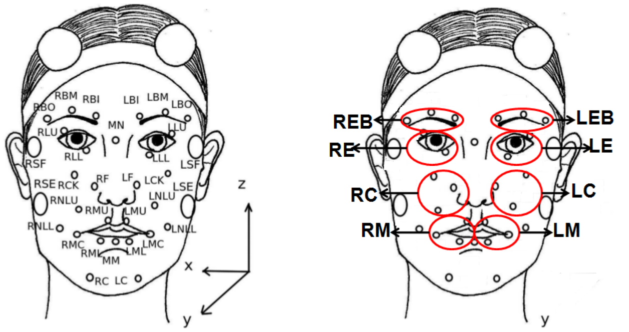


Fig. 1. Facial marker positions (left) and division of markers into the eight facial regions (right).

shapes and structures, and thus we can focus on purely expression related variability. After executing the above preprocessing steps, facial marker data from each subject for each expression is presented in the form of a matrix $\mathbf{D} \in \mathbb{R}^{\mathtt{T}\times\mathtt{M}}$ where $\mathtt{T}$ is the total number of time samples for an expression, and $\mathtt{M} = 28$ is the total number of facial markers. Note that in this work, we concentrate only on the horizontal and vertical displacement of the markers (x and z directions in Fig. 1).

We divide all facial expressions into two groups according to expressions produced by ASD and TD subjects. Within each group, the expressions were further partitioned into six emotion categories: Anger, Disgust, Fear, Happiness, Sadness and Surprise. ASD and TD subjects are analyzed and compared within each emotion category.

3.1. Dynamical Complexity Analysis

We begin with investigating the complexity of underlying mechanisms that generate facial expressions in children with and without ASD. We hypothesize that complexity will be lower for the ASD group. Traditional entropy measures assess the complexity of a system by quantifying local predictability or irregularity at a single scale and treat data from multiple variables as independent univariate systems [14]. Complexity analysis of a multivariate dynamic system requires the assessment of long-range linear/non-linear correlations within and across channels at multiple spatial and temporal scales. A recently developed entropy measure, namely the *multivariate multiscale entropy* (MMSE), [15, 16] is capable of quantifying the inherent complexity of a system by detecting dynamic structures or regularity within and across channels at multiple temporal scales.

Consider a multivariate time series $\mathbf{D}$ as above. For a given temporal scale factor ϵ, a coarse-grained version of $\mathbf{D}$ is obtained by partitioning each channel into $\mathtt{T}/\epsilon$ non-overlapping segments and averaging the values within each segment. Given a time lag vector $\tau = [\tau_1, \tau_2, ..., \tau_{\mathtt{M}}]$ and an embedding vector $\mathbf{m} = [m_1, m_2, ..., m_{\mathtt{M}}]$, all possible composite delay vectors are formed by concatenating m_i components from the i^{th} channel sampled at the rate of τ_i where $i = 1, 2, ..., \mathtt{M}$. Multivariate sample entropy is then computed for the coarse-grained time series in terms of the conditional probability of two composite vectors being close (in sense of a distance metric) in an $(m + 1)$ dimensional space, given that they are close in m dimensional space. For further details refer to [17, 15, 16].

For every emotion category, each expression matrix, $\mathbf{D}$, is subject to MMSE analysis at $\epsilon = 1, 2, ..., 5$; a single score is obtained for each ϵ. Mean MMSE scores for the ASD and TD groups are computed at ϵ, and results are presented in Fig. 2. In general, one multivariate time series is considered more complex than the other when it has higher entropy at the majority of temporal scales [16]. Results in Fig. 2 show that (i) TD group has a more complex expression generating mechanism than the ASD group for emotions like Disgust, Fear, Sadness and Surprise; (ii) For Sadness, the difference between the groups is the largest, indicating that expressions within this emotion group are likely to induce more atypicality to the observers; (iii) Sadness and Fear are more complex emotions compared to others; (iv) For Anger and Happiness, ASD and TD groups do not exhibit very clear differences in complexity.

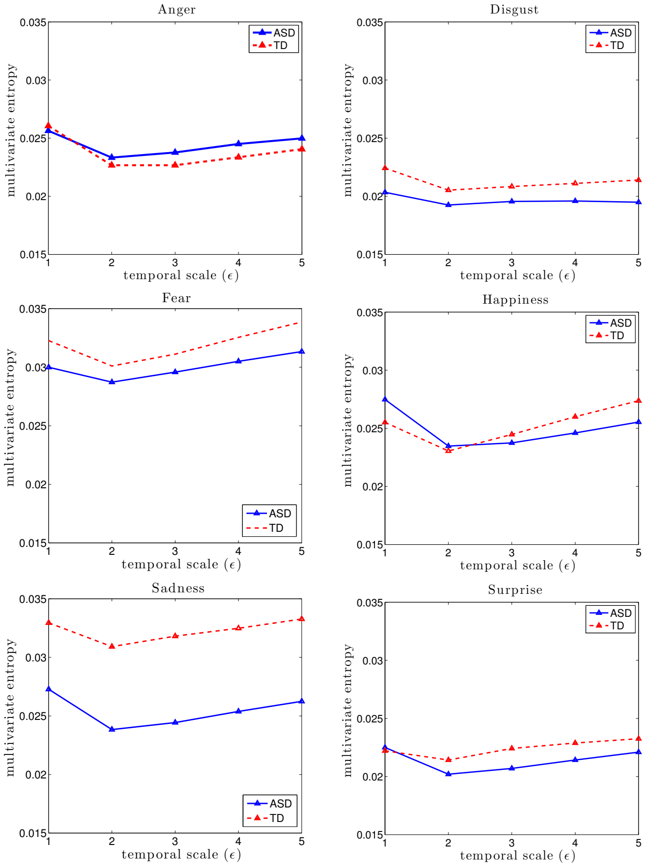


Fig. 2. Analysis of dynamical complexity computed in terms of multivariate entropy at multiple time scales for ASD and TD population for each emotion group.

3.2. Analysis Based on Local Regions

For robust processing and interpretability of facial behavior, we divide the markers into 8 regions as shown in Fig. 1, and perform analysis at the region level. These regions are: left eyebrow (LEB), right eyebrow (REB), left eye (LE), right eye (RE), left cheek (LC), right cheek (RC), left mouth (LM), and right mouth (RM). Note that only 22 markers are considered in the region-based analysis (unless mentioned otherwise), while all 28 markers are used during the complexity analysis.

3.2.1. Autoregressive Modeling

In this section, we build a reference model for each TD subject, and investigate how the temporal dynamics of ASD subjects diverge from the reference models within each emotion category.

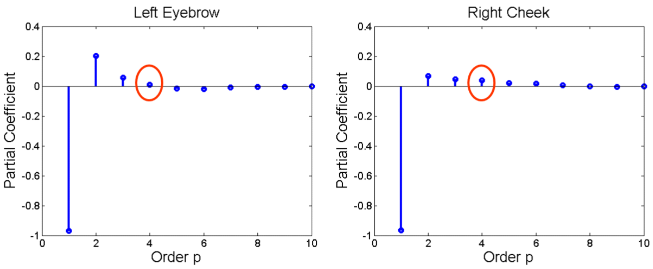


Fig. 3. Sample plots of partial autocorrelation coefficients as a function of order p for the LEB and RC regions. A similar trend is observed for all other regions.

To this end, we average the (x_t, y_t) coordinates of frame t over the markers within each region, and compute the L_2 distance using the averaged coordinates, i.e. $d_t = \sqrt{(\bar{x}_t^2 + \bar{y}_t^2)}$. This time series describes the dynamic evolution of an expression within each facial region. Autoregressive (AR)) models are popular for describing time-varying processes. Given a TD or ASD subject, we employ an AR model to capture the temporal dynamics of the representative time series of each facial region. An AR model of order p is defined as follows:

$$d_t = \sum_{i=1}^{p} \alpha_i d_{t-i} + \sigma_t, \qquad (1)$$

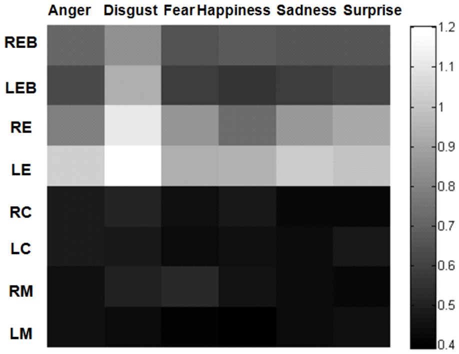
where σ_t is white noise, and $\{\alpha_i\}_{i=1}^{p}$ are the model parameters which parameterize the overall temporal dynamics of the given time series. Accordingly, the dynamics of the j^{th} facial region of the k^{th} TD or ASD subject are represented by a p-dimensional feature vector $\mathbf{f}_j^k = [\alpha_1, \alpha_2, \cdots, \alpha_p]$.

To determine the order, p, of the model, we examine the partial autocorrelation coefficients in relation to p for each facial region. By averaging the coefficients of each p across all the TD and ASD subjects, we find that the mean coefficient value converges at $p = 4$ for all the facial regions. Therefore, we use a 4^{th} order AR model for our analysis. Fig. 3 presents sample plots of the partial autocorrelation coefficients as a function of p for LEB and RC regions.

Within each emotion group, we compute the region-based distance between the dynamic feature vectors of each ASD-TD subject pair. Such a distance measures the dynamical divergence of ASD subjects from the reference (TD subjects) with respect to each facial region. Fig. 4(a) visualizes the mean region-based distance of facial dynamics between pairwise ASD and TD subjects in each emotion category. We can observe that the ASD subjects in the disgust category generally show the largest difference of facial dynamics from the reference. This result is consistent with the observation in Section 3.1 that higher complexity difference between ASD and TD groups exists for the Disgust expressions. In addition, the distance between ASD and TD subjects in the upper region including eyebrows and eyes is significantly larger compared to the lower region containing cheek and mouth. In particular, the highest dynamical divergence of ASD subjects from the reference is observed for eye regions. We summarize the mean distance of upper and lower regions within each emotion category in Fig. 4(b). These results indicate that the lower complexity of facial expressions of ASD subjects may result largely from the lower activation of their upper-face regions, especially the eye regions.

3.2.2. Activation Analysis

In this section, we study and compare facial expressions of the ASD and TD subjects in terms of activation of regions. *Activation* of a region can be understood as the intensity of movement the region undergoes during an expression. We first investigate how close the mimicry performances of the ASD and TD subjects are to the stimuli that were presented as references; we then move on to analyze in what ways their mimicry performances differ.



(a)

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Upper	0.79	1.02	0.76	0.73	0.79	0.80
Lower	0.47	0.48	0.45	0.44	0.43	0.44

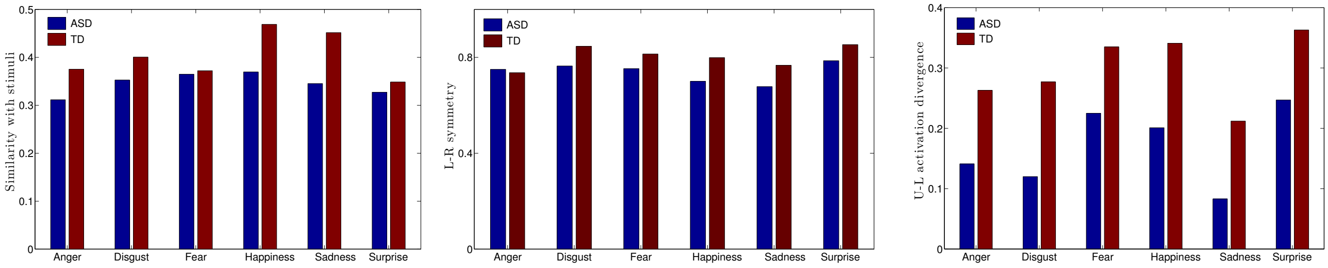
(b)

Fig. 4. Mean region-based distances between facial dynamics of ASD and TD subjects in each emotion group. Eye regions (RE and LE) show large differences between ASD and TD groups.

Table 1. Results of statistical t-tests for facial characteristics including all emotion categories with $N = 45$

Similarity with stimuli	
Correlations between computed activation and manual ratings	Lower correlations for ASD, $p = 0.024$
Left-Right activation symmetry	
Correlations between left and right regions	Lower correlations for ASD, $p = 0.0554$
Upper-Lower activation divergence	
Difference in activation between upper and lower regions	Lower divergence for ASD, $p = 5.23e$-4

Similarity with stimuli: In order to be able to compare the two groups with respect to the stimuli, we collected manual annotations *only* for the 36 reference stimuli videos - the clips the children were shown and instructed to mimic. Four experts rated each facial region based on how active each region appears during an expression. A score between 0 and 5 was given to each of the 8 facial regions (see Fig. 1) where a score of 0 indicates no activation, and 5 indicates high activation. These scores are averaged across raters to obtain average ratings per region per stimuli video. Each reference video is associated with a rating vector $\mathbf{r} \in \mathbb{R}^8$ containing average activation ratings for 8 regions. Annotations are available only for the stimuli videos. For the expressions of ASD and TD subjects, we compute a measure of activation from the facial marker data itself. This measure is expected to correspond with the activation perceived by the annotators. Intuitively, perception of activation of a region is associ-



(a) (b) (c)

Fig. 5. Comparison between ASD and TD subjects in terms of (a) Similarity with stimuli (b) Left-Right activation symmetry, and (c) Upper-Lower face activation divergence for individual emotion group.

ated with how much the region moves; hence we define activation as the total amount of motion exhibited by all the markers in that facial region over the duration of an expression.

Consider a facial region that contains $\mathtt{P}$ ($\subset$ $\mathtt{M}$) markers. Let the coordinates of the i^{th} marker be $(x_1^i, y_1^i), (x_2^i, y_2^i), ..., (x_\mathtt{T}^i, y_\mathtt{T}^i)$, where $i = 1, 2, .., \mathtt{M}$ and $\mathtt{T}$ is the total number of time samples. An activation score $\mathtt{a}$ for the region is computed as follows:

$$\mathtt{a} = \sum_{i=1}^{P} \frac{1}{\mathtt{T}} \sum_{t=1}^{\mathtt{T}} (|x_{t+1}^i - x_t^i| + |y_{t+1}^i - y_t^i|) \qquad (2)$$

After computing $\mathtt{a}$ for each local region, each data sample $\mathbf{D}$ is represented by a vector $\mathbf{a} \in \mathbb{R}^8$ containing activation values for all regions: $\mathbf{a} = [\mathtt{a}_{\mathtt{LEB}}, \mathtt{a}_{\mathtt{REB}}, \mathtt{a}_{\mathtt{LE}}, \mathtt{a}_{\mathtt{RE}}, \mathtt{a}_{\mathtt{LC}}, \mathtt{a}_{\mathtt{RC}}, \mathtt{a}_{\mathtt{LM}}, \mathtt{a}_{\mathtt{RM}}]$.

To study how well the ASD and TD subjects mimic a facial expression $\mathtt{E}$, we compute correlation between the computed activation $\mathbf{a}_\mathtt{E}$ and the manual annotations of the corresponding stimuli, $\mathbf{r}_\mathtt{E}$. Correlation coefficients are computed for each sample, and are averaged across all emotions for each subject in ASD and TD groups. A two-sample t-test is carried out with $N = 45$ (24 ASD + 21 TD). Significant difference between the groups is observed (see Table 1). Results for individual emotion category are presented in Fig. 5(a), which show that mimicry performance of the ASD group is less similar to the stimuli as compared to the TD group. The largest difference between groups is observed for Sadness and Happiness emotions indicating that ASD subjects have higher difficulty in mimicking these emotions. Whether this inferior mimicry performance of the ASD subjects is due difficulty in perception of the stimuli and/or in reproducing the gestures is an open question.

Left-Right (L-R) activation symmetry: Bilateral symmetry is an important characteristic of facial expressions. To measure this quantity, activation corresponding to the left and right sides of a face are computed as $\mathbf{a}_L = [\mathtt{a}_{\mathtt{LEB}}, \mathtt{a}_{\mathtt{LE}}, \mathtt{a}_{\mathtt{LC}}, \mathtt{a}_{\mathtt{LM}}]$ and $\mathbf{a}_R = [\mathtt{a}_{\mathtt{REB}}, \mathtt{a}_{\mathtt{RE}}, \mathtt{a}_{\mathtt{RC}}, \mathtt{a}_{\mathtt{RM}}]$. Left-Right activation symmetry for each data sample is measured in terms of correlation between $\mathbf{a}_L$ and $\mathbf{a}_R$. Correlation coefficients for ASD and TD subjects (averaged across all emotions for each subject) were used to perform a two-sample t-test with $N = 45$. Marginally significant differences are observed between the groups (Table 1). Results for individual emotion group are presented in Fig. 5(b), which show that ASD group has lower LR symmetry compared to the TD group; the difference is more pronounced for expressions of sadness. Despite subtle differences, most expressions are deemed symmetric, and lack of facial symmetry in ASD subjects may give rise to a sense of awkwardness.

Upper-Lower (U-L) activation divergence: Activation divergence between upper and lower regions signify the range of activation for an expression. Intuitively, this quantity is associated with how animated an expression is. The activation corresponding to the upper and lower regions of a face, $\mathtt{a}_{\mathtt{up}}$ and $\mathtt{a}_{\mathtt{lr}}$, are computed as follows:

$$\mathtt{a}_{\mathtt{up}} = \frac{\sum_{j \in \{\mathtt{LEB,REB,LE,RE}\}} \mathtt{a}_\mathtt{j}}{\sum_{i=1}^{\mathtt{M}} \mathtt{a}_\mathtt{i}}, \quad \mathtt{a}_{\mathtt{lr}} = \frac{\sum_{j \in \{\mathtt{LC,RC,LM,RM}\}} \mathtt{a}_\mathtt{j}}{\sum_{i=1}^{\mathtt{M}} \mathtt{a}_\mathtt{i}} \qquad (3)$$

For all ASD and TD subjects, activation in upper region is much less than that in lower region for all expressions. Activation divergence $(\mathtt{a}_{\mathtt{lr}} - \mathtt{a}_{\mathtt{up}})$, a positive quantity, is computed for each subject as before, and group difference is obtained. Significant difference is noted between the two groups (Table 1). Results for individual emotion categories are presented in Fig. 5(c), which show that ASD group has lower UL activation difference compared to the TD group; the difference is more pronounced for emotion category Disgust. This observation is consistent with the higher difference in upper and lower region facial dynamics obtained in time-series modeling for Disgust in Section 3.2.1. This is also suggestive of lower dynamic complexity of facial expressions in ASD.

4. CONCLUSION

In this paper, we analyzed facial expressions of children with ASD using MoCap data. The objective of this analysis is to quantify the emotional expressive atypicality often perceived by observers. We studied various global and local (region-based) characteristics using various signal processing and time series analysis tools.

Our major findings are: (i) overall, ASD subjects have less complex facial expressions, supporting the well known complexity loss theory in medical science under a disorder or disease [18]; (ii) the differences in facial dynamics between ASD and TD come mainly from the eye region; (iii) ASD subjects underperform at mimicking the stimuli, exhibit lower bilateral facial symmetry, and produce less variations across facial regions in terms of strength of activation; (iv) in general, group differences are found to be more pronounced for emotions with negatively valence, like Disgust and Sadness. This suggests that these emotions are likely to induce a higher perception of atypicality among the observers. Future work will be directed towards investigating the differences when subjects are at rest position before and after expressions, and jointly analyzing facial expressions with emotion in speech.

5. REFERENCES

[1] Giorgio Celani, Marco Walter Battacchi, and Letizia Arcidiacono, "The understanding of the emotional meaning of facial expressions in people with autism," *Journal of autism and developmental disorders*, vol. 29, no. 1, pp. 57–66, 1999.

[2] Nurit Yirmiya, Connie Kasari, Marian Sigman, and Peter Mundy, "Facial expressions of affect in autistic, mentally retarded and normal children," *Journal of Child Psychology and Psychiatry*, vol. 30, no. 5, pp. 725–735, 1989.

[3] Ruth B Grossman, N. Pitre, A. Schmid, and K. Hasty, "First impressions: Facial expressions and prosody signal asd status to naive observers," in *Intl. Meeting for Autism Research*, 2012.

[4] Hans Asperger and Uta Trans Frith, "'autistic psychopathy'in childhood.," 1991.

[5] Ashley de Marchena and Inge-Marie Eigsti, "Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency," *Autism research*, vol. 3, no. 6, pp. 311–322, 2010.

[6] Ruth B Grossman, Rhyannon H Bemis, Daniela Plesa Skwerer, and Helen Tager-Flusberg, "Lexical and affective prosody in children with high-functioning autism," *Journal of Speech, Language, and Hearing Research*, vol. 53, no. 3, pp. 778–793, 2010.

[7] Daniel Bone, Chi-Chun Lee, Matthew P Black, Marian E Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan, "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody," *Journal of Speech, Language, and Hearing Research*, 2014.

[8] Daniel Bone, Matthew P Black, Chi-Chun Lee, Marian E Williams, Pat Levitt, Sungbok Lee, and Shrikanth Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist.," in *INTERSPEECH*, 2012.

[9] Joshua John Diehl and Rhea Paul, "Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders," *Research in autism spectrum disorders*, vol. 6, no. 1, pp. 123–134, 2012.

[10] Zhaojun Yang and Shrikanth Narayanan, "Analysis of emotional effect on speech-body gesture interplay," in *Interspeech*, 2014.

[11] Angeliki Metallinou, Ruth B Grossman, and Shrikanth Narayanan, "Quantifying atypicality in affective facial expressions of children with autism spectrum disorders," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[12] "http://facelab.emerson.edu/".

[13] Simon Baron-Cohen, *Mind reading [: the interactive guide to emotions*, Jessica Kingsley Publishers, 2003.

[14] Holger Kantz and Thomas Schreiber, *Nonlinear time series analysis*, vol. 7, Cambridge university press, 2004.

[15] Mosabber Uddin Ahmed and Danilo P Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Physical Review E*, vol. 84, no. 6, pp. 061918, 2011.

[16] Mosabber Uddin Ahmed and Danilo P Mandic, "Multivariate multiscale entropy analysis," *Signal Processing Letters, IEEE*, vol. 19, no. 2, pp. 91–94, 2012.

[17] Madalena Costa, Ary L Goldberger, and C-K Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical review letters*, vol. 89, no. 6, pp. 068102, 2002.

[18] Ary L Goldbergera, C-K Penga, and Lewis A Lipsitzb, "What is physiologic complexity and how does it change with aging and disease?," *Neurobiology of Aging*, vol. 23, pp. 23–26, 2002.