QUANTIFYING ATYPICALITY IN AFFECTIVE FACIAL EXPRESSIONS OF CHILDREN WITH AUTISM SPECTRUM DISORDERS

Angeliki Metallinou[1], Ruth B. Grossman[2,3] and Shrikanth Narayanan[1]

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

metallin@usc.edu, ruth_grossman@emerson.edu, shri@sipi.usc.edu

ABSTRACT

We focus on the analysis, quantification and visualization of atypicality in affective facial expressions of children with High Functioning Autism (HFA). We examine facial Motion Capture data from typically developing (TD) children and children with HFA, using various statistical methods, including Functional Data Analysis, in order to quantify atypical expression characteristics and uncover patterns of expression evolution in the two populations. Our results show that children with HFA display higher asynchrony of motion between facial regions, more rough facial and head motion, and a larger range of facial region motion. Overall, subjects with HFA consistently display a wider variability in the expressive facial gestures that they employ. Our analysis demonstrates the utility of computational approaches for understanding behavioral data and brings new insights into the autism domain regarding the atypicality that is often associated with facial expressions of subjects with HFA.

Index Terms— affective facial expressions, motion capture, functional data analysis, autism spectrum disorders

1. INTRODUCTION

Facial expressions provide a window to internal emotional state and are key for successful communication and social integration. Individuals with High Functioning Autism (HFA), who have average intelligence and language skills, often struggle in social settings because of difficulty in interpreting [1] and producing [2, 3] facial expressions. Their expressions are often perceived as awkward or atypical by typically developing observers (TD) [4]. Although this perception of awkwardness is used as a clinically relevant measure, it does not shed light into the specific facial gestures that may have elicited that perception. This motivates the use of Motion Capture (MoCap) technology and the application of statistical methods like Functional Data Analysis (FDA, [15]), that allow us to capture, mathematically quantify and visualize atypical characteristics of facial gestures. This work is part of the emerging Behavioral Signal Processing (BSP) domain

that explores the role of engineering in furthering the understanding of human behavior [5].

Starting from these qualitative notions of atypicality, our goal is to derive quantitative descriptions of the characteristics of facial expressions using appropriate statistical analyses. Through these, we can discover differences between TD and HFA populations that may contribute to a perception of atypicality. The availability of detailed MoCap information enables quantifying overall aspects of facial gestures such as synchrony and smoothness of motion, that may affect the final expression quality. Dynamic aspects of facial expressions are of equal interest, and the use of FDA techniques such us functional PCA (fPCA) provides a mathematical framework to estimate important patterns of temporal variability and explore how such variability is employed by the two populations. Finally, given that children with HFA may display a wide variety of behaviors [6], it is important to understand child-specific expressive characteristics. The use of multidimensional scaling (MDS) addresses this point by providing a principled way to visualize differences of facial expression behavior across children.

Our work proposes the use of a variety of statistical approaches to uncover and interpret characteristics of behavioral data, and demonstrates their potential to bring new insights into the autism domain. According to our results, subjects with HFA are characterized on average by lower synchrony of movement between facial regions, more rough head and facial motion, and a larger range of facial region motion. Expression-specific analysis of smiles indicates that children with HFA display a larger variability in their smile evolution, and may display idiosyncratic facial gestures unrelated to the expression. Overall, children with HFA consistently display a wider variability of facial behaviors compared to their TD counterparts, which corroborates with existing psychological research [6]. Those results shed light into the nature of expression atypicality and certain findings, e.g., asynchrony, may suggest an underlying impairment in the facial expression production mechanism that is worth further investigation.

2. RELATED WORK

Since early psychological works [7, 8], autism spectrum disorders (ASD) have been linked to the production of atypical facial expressions and prosody. Autism researchers have reported that the facial expressions of subjects with ASD are often perceived as different and awkward [2, 3, 4]. Researchers have also reported atypicality with the synchronization of expressive cues, e.g., verbal language and body gestures [9]. Inspired by these observations of asynchrony, we examine synchronization properties of minute facial gestures, which are often hard to describe by visual inspection.

Recent computational work aims to bring new understanding of this psychological condition and develop technological tools to help ASD individuals and psychology practitioners. Work in [10] describes eye tracking glasses to be worn by the practitioner and track gaze patterns of children with ASD during therapy sessions, while [11] introduces an expressive virtual agent that is designed to interact with children with ASD. Computational analyses have mostly focused on atypical prosody, where certain prosodic properties of subjects with ASD are shown to correlate with the severity of autism diagnosis [12, 13, 14]. In contrast, computational analysis of atypical facial expressions is a relatively unexplored topic.

Our analysis relies heavily on FDA methods, which were introduced in [15] as a collection of statistical methods for exploring patterns in time series data. A fundamental difference between FDA and other statistical methods is the representation of time series data as functions rather than multivariate vectors, which exploits their dynamic nature. FDA techniques have been successfully applied for quantifying prosodic variability in speech accents [16], and the analysis of tendon injuries using human gait MoCap data [17].

3. DATABASE

We analyze data from 37 children (21 with HFA, 16 TD) aged 9-14, while they perform mimicry of emotional facial expression videos from the Mind Reading CD, a common psychology resource [18]. Expressions last a few seconds and cover a variety of emotions including happiness (smile), anger (frown) and emotional transitions, e.g. surprise followed by happiness (mouth opening and smile). Children are instructed to watch and mimic those expressions. There are two predefined sets of expressions with 18 expressions each, covering similar expressions. Each child mimics the 18 expressions of one set. Children wear 32 facial markers, as in Figure 1, and are recorded by 6 MoCap cameras at 100 fps.

4. DATA PREPARATION

Four stability markers were used to factor out head movement (depicted in Fig 1 as larger markers in the forehead and ears).
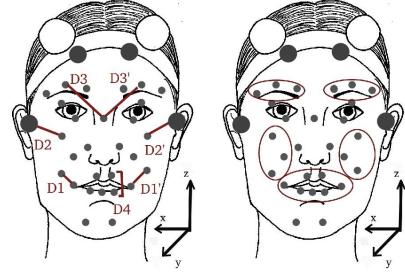

Fig. 1. Placement of facial markers and definition of facial distances (left) and facial regions (right).

The positions of the remaining 28 markers are computed with respect to the stability markers and used for further facial expression processing, while the the stability markers are used to provide head motion information. Facial data were further rotated to align with the (x,y,z) axes as depicted in Figure 1, and were centered to the origin of the coordinate system. Data visualization tools were developed to visually inspect the MoCap sequence and correct any artifacts.

We perform face normalization to smooth out subject-specific variability due to different facial structure, and focus on expression related variability. We apply the normalization approach proposed in our previous work [19], where each subject's mean marker positions are shifted to match to the global mean positions computed across all subjects. Finally, marker trajectories were interpolated to fill in gaps shorter than 1sec, that result from temporarily missing or occluded markers. We use cubic Hermite spline interpolation, which we empirically found to produce visually smooth results.

Marker data are then transformed into functional data. This process consists of approximating each marker coordinate time series, e.g., $x_1, ... x_T$, by a function say $\hat{x}(t) = \sum_{k=1}^{K} c_k \phi_k(t)$, where $\phi_k, k = 1, \dots, K$ are the basis functions and $c_1, c_2, \dots, c_k$ are the coefficients of the expansion. Conversion into functional data performs smoothing of the original time series, enables computing smooth approximations of high order derivatives of marker trajectories (Section 5), and enables FDA methods such as fPCA (Section 6).

We use B-splines as basis functions, which are commonly used in FDA because of their flexibility to model non-periodic series [15, 16]. Fitting is done by minimizing:

$$F = \sum_i [x_i - \hat{x}(t_i)]^2 + \lambda \int [D^2 \hat{x}(t)]^2 dt$$

where D^2 denotes second derivative and parameter λ controls the amount of smoothing (second term) relative to the goodness of fit of function $\hat{x}$ (first term). We choose $\lambda = 1$ empirically according to the Generalized Cross Validation

Table 1. Results of Statistical Tests of Global Facial Characteristics (t-test, difference of means)

comparison	result
Left-Right Face Synchrony	
left-right mouth corner correlations	Lower correlations for HFA, p=0.02
left-right cheek correlations	Lower correlations for HFA, p=0.07
left-right eyebrow correlations	Lower correlations for HFA, p=0.01
Upper-Lower Face Synchrony	
right eyebrow & mouth opening correlations.	Lower correlations for HFA, p=0.05
left eyebrow & mouth opening correlations	Lower correlations for HFA, p=0.03
Facial Motion Roughness ($i = 2$)	
mouth roughness	Higher roughness for HFA, p=0.02
right cheek roughness	Higher roughness for HFA, p=0.01
left cheek roughness	no difference
right eyebrow roughness	Higher roughness for HFA, p=0.07
left eyebrow roughness	no difference
Head Motion Roughness ($i = 2$)	
head roughness	Higher roughness for HFA, $p \approx 0$
Facial Motion Range	
upper mouth motion range	Higher range for HFA, $p \approx 0$
lower mouth motion range	Higher range for HFA, $p \approx 0$
right cheek motion range	Higher range for HFA, $p \approx 0$
left cheek motion range	no difference
right eyebrow motion range	no difference
left eyebrow motion range	no difference

(GCV) criterion [20]. The FDA analysis throughout the paper is performed using the FDA toolbox [20].

5. ANALYSIS OF GLOBAL CHARACTERISTICS OF AFFECTIVE EXPRESSIONS

We group the expressions into two groups containing expressions produced by subjects with TD and HFA, and perform statistical analysis of expressive differences. We examine properties inspired from psychology, such as synchrony [9], or properties that intuitively seem to affect the quality of the final expression, e.g., facial motion smoothness and range of marker motion (large range may suggest exaggerated expressions).

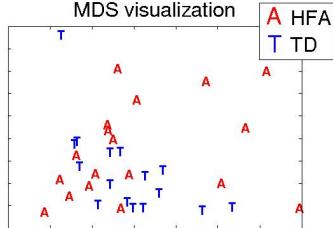TD and HFA groups contain roughly 16×18 and 21×18 expressions respectively, although certain samples are removed because of missing or noisy markers. When grouping together various facial expressions, we want to smooth out expression-related variability and focus on subject-related variability. Therefore, all metrics described below are normalized by mean shifting such that the mean of each metric per expression (and across multiple subjects) is the same across expressions.

We examined synchrony of movement across left-right and upper-lower face regions. For left-right comparisons, we measured facial distances associated with muscle movements, specifically mouth corner, cheek raising, and eyebrow raising. These distances are depicted as $D1, D2$ and $D3$ respectively for the right face and $D1', D2', D3'$ for the left face, in Fig 1. To measure their motion synchrony, we computed Pearson's correlation between $D1$-$D1'$, $D2$-$D2'$ and $D3$-$D3'$. For upper-lower comparisons, we measured mouth opening ($D4$, Figure 1) and eyebrow raising distances ($D3$ and $D3'$), and we computed Pearson's correlation between $D4$-$D3$, and $D4$-$D3'$. We examine the statistical significance of group differences in correlation using a difference of means t-test. Results are presented in Table 1 indicating lower facial movement synchrony for subjects with HFA.
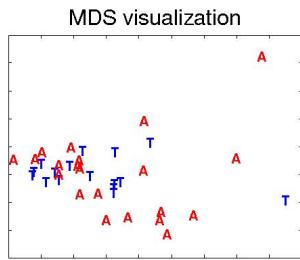
Although statistical tests reveal global differences, subject-specific characteristics are also of interest. Subject characteristics are visualized using multidimensional scaling (MDS), which is a collection of methods for visualizing the proximity of multidimensional data points [21]. MDS takes as input a distance (dissimilarity) matrix, where dissimilar data points are more distant, and provides methods for transforming data points into lower dimensions, typically two dimensions for easy visualization, while optimally preserving their dissimilarity.

Here, a data point is a subject associated with a multidimensional feature of average correlations. A subject performs 18 expressions, and for each we compute the 3 left-right facial correlations mentioned above. By averaging over expressions we compute a 3D average correlation feature per subject. Dissimilarity between subjects is computed by taking euclidean distances of their respective features. MDS uses this dissimilarity matrix to compute the distances between subjects in 2D space. Figure 2(a) shows the MDS visualization for TD subjects represented in blue 'T', and subjects with HFA in red 'A', when the average left-right symmetry correlations are used as features (we applied non-classical multidimensional scaling using the metricstress criterion, and confirmed that original dissimilarities are adequately preserved in 2D [21]). Intuitively, Figure 2(a) illustrates similarities of subjects with respect to left-right facial synchrony behavior. We notice that subjects with HFA generally display larger behavioral variability, although there is one TD outlier. This visualization could help clinicians understand subject-specific characteristics with respect to particular facial behaviors.

Smoothness of motion was investigated by computing

(a) Left-right face synchrony.

(b) Face region motion roughness.

Fig. 2. MDS visualization of similarities across subjects for left-right synchrony and facial region motion roughness metrics.

higher order derivatives of facial and head motion. We examine 5 facial regions, i.e., left and right eyebrows, left and right cheeks, and mouth (Figure 1). Each region centroid is computed by averaging the markers in that region, and for each centroid motion we estimate the absolute derivatives of order $i, i = 1, 2, 3$ averaged during the expression. We call this a roughness measure of order i. Similar computations are performed for head motion, where the head centroid is the average of the 4 stability markers. The results in Table 1, indicate more roughness of head and lower/right facial region motion for subjects with HFA. For lack of space we present only acceleration results ($i = 2$), but other order derivatives follow similar patterns. We perform MDS analysis using as features the average acceleration roughness measures per subject from mouth, right cheek and right eyebrow regions. We select those regions since, according to Table 1, they have significantly higher roughness measures. According to the resulting MDS visualization of Figure 2(b), subjects with HFA are more likely to be outliers and display larger variability.

Finally, we examine the range that facial regions traverse during an expression, i.e. range of motion for eyebrows, cheeks, upper and lower mouth region centroids. This range is significantly higher for the HFA group for lower and right face regions, as seen in Table 1.

6. QUANTIFYING EXPRESSION-SPECIFIC ATYPICALITY THROUGH FPCA

While previous analyses looked at global expression properties, here we perform expression-specific analysis focusing on dynamic expression evolution. As an example, we choose an expression of happiness, consisting of two consecutive smiles. This expression belongs to one of the two expression sets mentioned in Section 3, so it is only performed by 19 children (7 TD, 12 HFA), out of the 37 children in our database. Smiles are chosen for study because they are common in a variety of social interactions. Moreover, since this expression contains a transition between two smiles, it has increased complexity and is a good candidate for revealing typical and atypical variability patterns. Finally, mouth region expressions seem to differ between TD and HFA groups, as shown in Section 5, while the onset and apex of those expressions are easily detected by looking at mouth distances, for example distance $D1$ from Figure 1.

In Fig 3(a) we depict the mouth corner distance $D1$ during the two smiles expression from a TD subject (blue solid line), and a subject with HFA (red dashed line). The expression of the TD subject depicts a typical evolution where the 2 distance minima (black circles) correspond to the apex of the two smiles and are surrounded by 3 maxima (black squares), representing the beginning, middle (between smiles) and end of the expression respectively. The expression of the subject with HFA follows a more atypical evolution and contains seemingly unrelated motion, for example the oscillatory motion in the middle.

For better comparison, expressions are aligned such that the smile apices of different subjects coincide. We apply a method called landmark registration [15, 16], which uses a set of predefined landmarks, e.g., comparable events during an expression, and computes warping functions such that the landmark points of different expression realizations coincide. Here we define 5 landmarks; the 2 minima (smile apices) and 3 maxima described above. Landmarks are automatically detected by searching for local maxima/minima and are manually corrected if needed. In Figure 3(e) we show the subjects' expressions after landmark registration, where the two smile events are aligned and clearly visible.

After alignment, we compute principal components of expression variability using fPCA. fPCA is an extension of ordinary PCA that operates on a set of functional input curves, i.e., $x_i(t), i = 1, .., N$ (here N=19). fPCA iteratively computes eigenfunctions $\xi_j(t)$ such that the data variance along the eigenfunction is maximized at each step j, i.e., maximize $Var\big(\int \xi_j(t)x_i(t)dt\big)$, subject to normalization and orthogonality constraints, i.e., $\int \xi_j^2(t)dt = 1$ and $\int \xi_j(t)\xi_k(t)dt = 0, \forall k \neq j$. The resulting set of eigenfunctions represents an orthonormal basis system where input curves are decomposed into principal components of variability. The PCA score of input curve $x_i(t)$ along $\xi_j(t)$ is defined as $c_{ij} = \int \xi_j(t)x_i(t)dt$
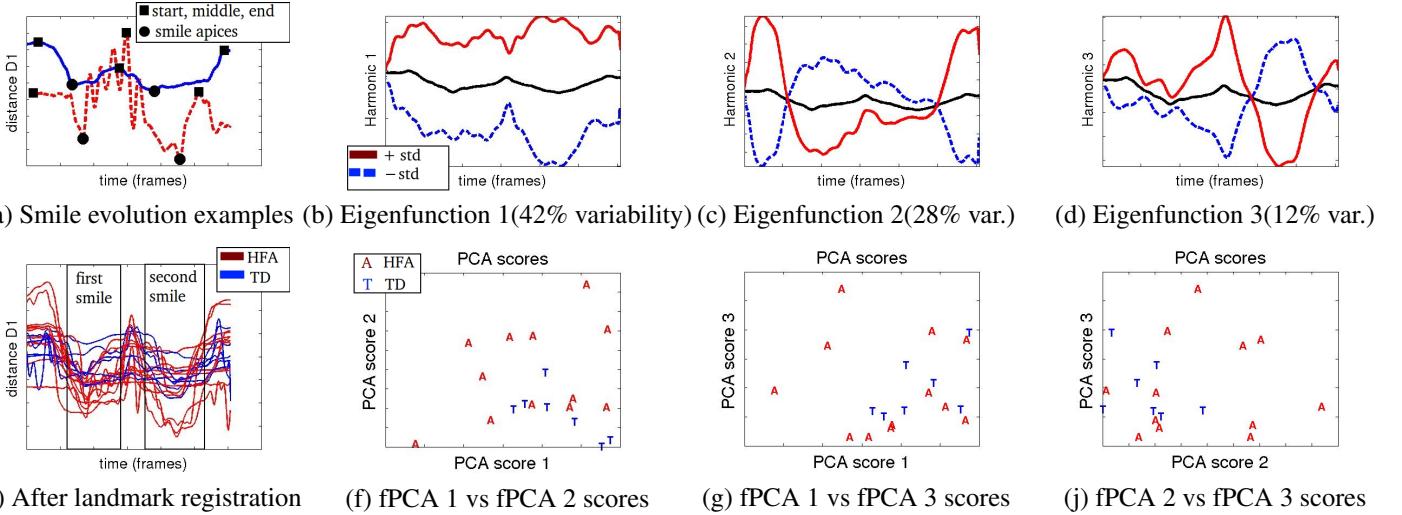
(a) Smile evolution examples (b) Eigenfunction 1(42% variability) (c) Eigenfunction 2(28% var.) (d) Eigenfunction 3(12% var.)

(e) After landmark registration (f) fPCA 1 vs fPCA 2 scores (g) fPCA 1 vs fPCA 3 scores (j) fPCA 2 vs fPCA 3 scores

Fig. 3. Analysis of the expression of two consecutive smiles. Plots of distance $D1$ from multiple subjects before and after landmark registration (subjects with HFA in red, TD in blue lines). Plots of the first 3 fPCA harmonics and scatterplots of corresponding fPCA scores.

(assuming mean subtracted curves for simplicity).

Figures 3(b)-(d) present the first 3 eigenfunctions (harmonics), cumulatively covering 82% of total variability. As in [15], the black line is the mean curve, and the solid red line and dashed blue lines illustrate the effect of $\xi_j(t)$ by adding or subtracting respectively $std(c_j) \times \xi_j(t)$ to the mean curve (standard deviation $std(c_j)$ is computed over all PCA scores c_{ij}). Although, these three first principal components of variability are estimated in a data-driven way, they seem visually interpretable. They respectively account for variability of: overall expression amplitude (Fig. 3(b)), the smile width as defined by the curve dip happening between initial and ending points (Fig. 3(c)), as well as mouth closing and second smile apex happening in the second half of the expression (Fig. 3(d)). Figures 3(f)-(j) present scatterplots of the first 3 PCA scores of different subjects' expression. Subjects with HFA display a wider variability of PCA score distribution, which translates to wider variance in the expression evolution, and potentially contributes to an impression of atypicality. Note that fPCA is unaware of the diagnosis label when it decomposes each expression into eigenfunctions. This decomposition naturally reveals differences between the TD and HFA groups in the way the smile expression evolves.

We performed fPCA analysis of various expressions, and made similar observations of greater variability in the evolution of expressions by subjects with HFA, mostly for complex expressions containing transitions between facial gestures e.g., mouth opening and then smile, consecutive smiles of increasing width. Mimicry of such expressions might be challenging for children with HFA, and may thus reveal differences between expressions of the two groups. Note that since we analyze posed expressions from children we would expect some degree of unnaturalness from the subjects. However, Figures 3(a) and (f)-(j) suggest a different nature and wider variance of expressive choices produced by subjects with HFA, which are sometimes unrelated to a particular expression. For example, the oscillatory motion displayed by the subject with HFA at Figure 3(a) appears in other facial regions and other expressions of the same subject, and seems to be an idiosyncratic facial gesture.

7. CONCLUSIONS AND FUTURE WORK

We have focused on quantifying atypicality in affective facial expressions, through the statistical analysis of MoCap data from facial gestures, which are hard to describe quantitatively by visual inspection. For this purpose, we have demonstrated the use of various data representation, analysis and visualization methods for behavioral data. We have found statistically significant differences in the affective facial expression characteristics between TD children and children with HFA. Specifically, children with HFA display more asynchrony of motion between facial regions, more head motion roughness, and more facial motion roughness and range for the lower face regions, compared to TD children. Children with HFA also display a wider variability in the expressive choices that they employ. Our results shed light on the characteristics of facial expressions of children with HFA and support qualitative psychological observations regarding atypicality of those expressions [2, 3]. In general, our described analyses could be applied in various time series data of human behavior, where the main goal is the discovery and interpretation of data patterns.

Our future work includes analysis of a wider range of

expressions including both positive and negative emotions. We also plan to obtain perceptual measurements of awkwardness of the collected expressions by TD observers in order to further interpret our findings. Finally, the availability of facial marker data enables analysis-by-synthesis approaches, including facial expression animation using virtual characters. It would be interesting to explore if the development of expressive virtual characters for animation and manipulation of typical and atypical facial gestures could provide any further insights to autism practitioners, or provide a useful educational tool of facial expression visualization for individuals with HFA.

8. REFERENCES

[1] G. Celani, M. W. Battacchi, and L. Arcidiacono, "The understanding of the emotional meaning of facial expressions in people with autism," *Journal of Autism and Developmental Disorders*, vol. 29, pp. 57–66, 1999.

[2] R. B. Grossman, L. Edelson, and H. Tager-Flusberg(in press)., "Production of emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism.," *Journal of Speech Language & Hearing Research, in press*.

[3] N. Yirmiya, C. Kasari, M.Sigman, and P. Mundy, "Facial expressions of affect in autistic, mentally retarded and normal children," *Journal of Child Psychology and Psychiatry*, vol. 30, pp. 725–735, 1989.

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

[5] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proc. of the IEEE*, to appear 2012.

[6] L. D. Shriberg, R. Paul, J. L. McSweeny, A. M. Klin, D. J.Cohen, and F. R. Volkmar, "Speech and prosody characteristics of adolescents and adults with high-functioning autism and asperger syndrome.," *Journal of Speech Language & Hearing Research*, vol. 44, pp. 1097–1115, 2001.

[7] H. Asperger, *Autistic psychopathy in childhood*, Autism and Aspergers Syndrome. Cambridge: Cambridge University Press., 1944.

[8] L. Kanner, "Autistic disturbances of affective contact," *Acta Paedopsychiatr*, vol. 35, pp. 100–136, 1968.

[9] A. deMarchena and I.M. Eigsti, "Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency," *Autism Research*, vol. 3, pp. 311–322, 2010.

[10] Z. Ye, Y. Li, A. Fathi, Y. Han, A. Rozga, G. D. Abowd, and J. M. Rehg., "Detecting eye contact using wearable eye-tracking glasses," in *Proc. of 2nd Intl. Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (PETMEI) in conjunction with UbiComp*, 2012.

[11] E. Mower, M. Black, E. Flores, M. Williams, and S. S. Narayanan, "Design of an emotionally targeted interactive agent for children with autism," in *Proc. of IEEE Intl. Conf. on Multimedia & Expo (ICME)*, 2011.

[12] J. J. Diehl and R. Paul, "Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders," *Research in Autism Spectrum Disorders*, vol. 1, pp. 123–134, 2012.

[13] J.P.H. Van Santen, E.T. Prud'hommeaux, L.M. Black, and M. Mitchell, "Computational prosodic markers for autism," *Autism*, vol. 14, pp. 215–236, 2010.

[14] D. Bone, M. P. Black, C.-C. Lee, M. E. Williams, P. Levitt, S. Lee, and S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in *Proc. of Interspeech*, 2012.

[15] J.O. Ramsay and B.W. Silverman, *Functional data analysis, 2nd ed.*, New York, Springer, 2005.

[16] M. Gubian, L. Boves, and F. Cangemi, "Joint analysis of f0 and speech rate with functional data analysis," in *Proc. of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011.

[17] O.A. Donoghue, A.J. Harrison, N. Coffey, and K. Hayes, "Functional data analysis of running kinematics in chronic achilles tendon injury.," *Medicine & Science in Sports and Exercise*, vol. 40, pp. 1323–1335, 2008.

[18] S. Baron-Cohen and J. Kingsley, *Mind Reading: The Interactive Guide to Emotions*, London, UK, 2003.

[19] A. Metallinou, C. Busso, S. Lee, and S. Narayanan, "Visual emotion recognition using compact facial representations and viseme information," in *Proc. of ICASSP*, 2010.

[20] J.O. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and Matlab*, vol. 66, Blackwell Publishing Inc, 2010.

[21] Trevor F. Cox and Michael A. Cox, *Multidimensional Scaling. 2nd ed.*, Chapman & Hall/CRC, 2000.