A Wearable Camera Detects Gaze Peculiarities during Social Interactions in Young Children with Pervasive Developmental Disorders

Silvia Magrelli, Basilio Noris, Patrick Jermann, François Ansermet, François Hentsch, Jacqueline Nadel, and Aude G. Billard Member, IEEE

Abstract—We report on the study of gazed, conducted on children with pervasive developmental disorders (PDD), by using a novel head-mounted eye-tracking device called the WearCam. Due to the portable nature of the WearCam, we are able to monitor naturalistic interactions between the children and adults. The study involved a group of 3 to 11 year-old children (n=13) with PDD compared to a group of typically developing (TD) children (n=13) between 2 and 6-years old. We found significant differences between the two groups, in terms of the proportion and the frequency of episodes of directly looking at faces during the whole set of experiments.

We also conducted a differentiated analysis, in two social conditions, of the gaze patterns directed to an adult’s face when the adult addressed the child either verbally or through facial expression of emotion. We observe that children with PDD show a marked tendency to look more at the face of the adult when she makes facial expressions rather than when she speaks.

Index Terms—Pervasive Developmental Disorders, Eyetracking, Atypical Gaze to Human Faces, Visual Perception, Social Attention, Facial Expressions, Social Sounds

I. INTRODUCTION

Pervasive developmental disorders (PDD) (DSM-IV-TR [1]) cover a large range of pervasive developmental impairments that affect an individual’s way of communicating and interacting socially. Autistic disorder, the most severe variation of pervasive developmental disorders, is characterized by severe abnormalities in social relationships, in language and in reactions to the environment. As stated by Johnson et al. [2], the term Autism Spectrum Disorder (ASD) was introduced “to reflect the broader spectrum of clinical characteristics that now define autism. ASDs represent three of the pervasive developmental disorders (PDD) defined in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV [3]), and the newer Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR [1]).

Autistic disorder (AD), Asperger syndrome (AS [this terminology will be used in this report, although “Asperger’s disorder” is used in the aforementioned publications]), and pervasive developmental disorder-not otherwise specified (PDD-NOS).” Included under PDD (DSM-IV-TR [1]) are also Rett’s disorder and childhood disintegrative disorder.

Over the past several decades, research in ASD has stressed social attention as a cardinal framework of measures for defining the nature of deficits in social skills [4]. Social attention impairments include a lack of interest in faces, abnormal social orienting and joint attention, atypical eye-to-eye contact and insensitivity towards emotional expressions. These features are considered key characteristics of ASD and are used for diagnosis of children at 24 months of age [1].

As observed by Filipek et al. [5], retrospective analyses of home videotapes or parental descriptions of children’s behavior within the first two years of life prior to diagnosis ( [6], [7]) lead to the identification of peculiarities in social attention in early infancy; these particularities distinguish ASD from other developmental disabilities. These studies are important as they provide a detailed account of the behavior of children with ASD in real-life settings where manifestations of social impairments are more evident [8]. As suggested by Čeponiené et al. [9], the complexity and fast changing nature of social stimuli during natural interactions can exacerbate the impaired sensory representation and processing in ASD ( [10], [11], [12], [13]). Moreover, Birmingham et al. [14] propose that the analysis of spontaneous responses in complex social environments could capture the fundamental mechanism underlying social attention in ASD. In this study, we set forth to analyze the gaze strategies of children with PDD, while they are engaged in naturalistic interactions with an adult.

The literature concerning childhood disintegrative disorder pervasive and Rett’s disorder is scarcer when compared to the research effort dedicated to the study of autism spectrum disorders (Malhotra and Gupta [15], Matson and Nebel-Schwalm [16]). Several studies compare the behavior of individuals with PDD with the results reported in the literature on ASD (van der Geest et al. [17], Chakrabarti and Fombonne [18], Herring et al. [19], Volkmar et al. [20]). Thus, mainly, we refer to the research on ASD and, when possible, to the results related to individuals with PDD.

There are mixed findings about the impairments in multimodal integration in ASD (see Foxe et al. [21] for a review).
However, individuals affected by ASD seem to have major difficulties when detecting intermodal correspondence between facial and vocal information (222), (223), (224)). Moreover, Magnée et al. (25) found impairments in audio-visual integration of speech stimuli in male adults with PDD. Recently, Silverman et al. (26) examined iconic gesture comprehension in adolescents with high functioning autism and found that the subjects with diagnosis present difficulties in simultaneously processing speech and gestures.

To determine signs of atypical attention to faces and the dependency of this attention on the type of social cues expressed by the face, we contrast gaze strategies when the face of the person either makes facial expressions or speaks. In addition, we provide a comparison with the gaze patterns directed to the toys used during the interactions. Facial expressions of emotion are mono-modal social stimuli conveyed solely through visual information (the emotion of the eye, the position of the mouth, etc), whereas, speaking involves two modes (27): It provides a visual stimulus (movements of the mouth, the intensity of the gaze while the person is speaking, etc.) combined with an auditory stimulus (the sound and the tone of the speech, etc.). As mentioned above, there is ample evidence that social impairments in PDD are exacerbated by social demands. Hence, we could expect a differentiated response to the types of social events considered here, i.e., we expect less time spent looking at faces when the adult addresses the child verbally compared to when the adult is silent but her face expresses an emotion.

Much research has been devoted to studying the neural and behavioural components of impairments in overt attention to faces. Although people with ASD normally detect gaze direction (28), visual attention peculiarities have been extensively reported. Atypical patterns of eye contact and inattention to faces have been reported in toddlers (29) and in pre-adolescents and adolescents (30) with ASD. In children with a high risk of autism (later diagnosed at 24 months), reduced attention to faces has been observed by retrospective analysis of home videos taken during the first 12 months of these children’s lives (31), (32). Moreover, several studies (33), (34), (35) point out the tendency of individuals with ASD to look at people less than objects and for shorter durations.

In recent years, several portable eye-tracking systems have been developed; they allow for a more natural recording of the child’s vision (36). Comparative studies relying on eye-tracking methods analyze gaze exploration in static and dynamic recordings. Measurements of the duration and region of gaze show that, for example, people with ASD tend to look predominantly at the mouth, whereas control subjects look predominantly at the eyes with both neutral and emotional faces (37), (38), (39), (40)). Jones and Klin (41) conducted a longitudinal eye-tracking study of children viewing naturalistic interactions. They report that the attention to the eyes, of children with ASD, begins to decline as early as 2 months of age. However, they do not observe any statistical difference in fixations to the mouth between children with ASD and TD children from 2 to 24 months of age.

In general, there are contradictory reports on people with ASD and their difficulties in identifying and processing information from faces in general (see Simmons et al. [8] for a comprehensive survey) and from facial expressions in particular (see Harms et al. [42] for a review). These discrepancies can be explained by other factors, such as methodological differences in task demands and the inherent diversity of the autistic spectrum disorders (42), (43)). For instance, Speer et al. (44) report in their study that children with autism, ages 9 to 18, differed from their typically developing peers in their gaze behaviors towards dynamic videos of social scenes. However, the authors report that the two groups of children showed similar gaze patterns to faces when showed in static photos.

Visual social-attention impairment in people with ASD cannot be related exclusively to either visual-feature processing or low motivation in social involvement (8). For Mundy et al. (45), (46), there exists an interaction between the early manifestations of social inattention, which does not allow children with ASD to gather social information during infancy, and the disruption of normal brain and behavioral development. Senju and Johnson (47), relying on functional imaging results, hypothesize that perceived eye-contact (which they term eye contact effect) modulates the activation of the social brain network via a subcortical route toward cortical areas involved in social cognition. In a more recent investigation, Senju et al. (48) find that people with autism present normal functionalities of the subcortical route for the generation of saccades. Whereas, they observe that, when people are looking at images of upright faces, subcortical control is involved in TD people, but not in people with ASD. This result suggests that the subcortical route might not be developed for face processing and thus supports similar findings (49), (50), (47)). These studies emphasize the key role played by the interaction between the subcortical and the cortical structures on social cognition development in people with autism.

Several retrospective studies report that children with ASD have severe social attention impairments when complex social stimuli (e.g., both visual and auditory) were presented them in naturalistic settings. For instance, Baranek et al. (11) compare the behaviors of infants with autism between the ages of 9 to 12-months and those of mentally retarded and typically developed peers. They find that people with ASD show poor visual attention and aversion to social touch and they respond less often to their names. Werner et al. (51) and Dawson et al. (52) report on impaired spontaneous attention shifting (i.e., social orienting) to social stimuli (e.g., calling their names and clapping of hands) in 10-month old children later diagnosed with ASD. In toddlers with autism, Dawson et al. (53) also find social orienting impairment. They observe that people with autism orient to social sound less than to non-social sounds. Recently, Chawarska et al. (54) conducted a semi-naturalistic eye-tracking study that involved children with ASD (n = 54, mean age = 1.8 years), TD children (n = 48, mean age = 1.7 years) and children with developmental delays (n = 20, mean age = 1.7 years), matched on verbal and non-verbal mental age. Chawarska et al. compared the gaze behaviours towards faces. The experimental tasks consisted in showing to the children video recordings with increasingly complex social information. Chawarska et al. find that children with ASD spent a significantly lower proportion of time looking...
at the face than the other groups did. This effect was most
evident in the dynamic scenes consisting in dyadic bids and
joint attention, i.e. when the actor in the recorded video spoke
or looked directly at the camera as if she were interacting with
the child. The authors hypothesize that the differences across
groups could be explained by the degree of social complexity
involved in the scenes, e.g., in the presence of social cues
such as eye contact and speech. In another recent eye-tracking
study, Shic et al. [55] find that 6-month-old children, who were
later diagnosed with ASD, looked proportionally less time at
the inner features of the face when an actor was speaking,
compared to all other groups.

Abnormal processing of speech sounds appears as a serious
deficit in both children and adults with ASD and might be the
basis of many of the social impairments in children with ASD
([56], [57]). This impairment in the processing of sounds
seems to pertain solely to speech sound, and not to other
sounds ([58], [9], [59], [60], [61]). There is mixed evidence of
neural correlates to these deficits in processing speech sounds.
Boddaert et al. ([56]) find abnormal auditory processing in
speech related areas of the superior temporal sulcus, STS, ([56], [62]) associated with ASD. However, Gervais et al. ([62])
show that the activation of the STS voice-selective areas, in
response to non-vocal sounds, is intact in autism spectrum
disorders.

In our study, we used a novel, light and unobtrusive eye-
tracker, the WearCam ([63]). This device overcomes one of
the major disadvantages of other eye-tracking systems: By
allowing the recordings of a broad region of the child’s field of
view, it can provide us a comprehensive understanding of the
child’s visual attention focus. This advantage, together with
the non-invasiveness of the device, enables the use of our eye
tracker in real-life conditions and contrasts with the request for
constrained experimental conditions by the other eye-trackers.
Therefore, the WearCam technology offers us the possibility
to monitor visual attention in naturalistic settings. The WearCam
device also enables us to record the gaze separately from the
head direction. Smith et al. attached a small camera to
the forehead of typically developing children during a dyadic
naturalistic interaction ([64], [65]). This provides a first person
point of view but does not give information about the child’s
gaze, as the view it transmits is aligned with the head rather
than with the eyes. When studying gazes, monitoring eye
motion separately from head motion offers a unique means for
quantifying measures such as the frequency and the duration of
events of fixations. The presence of particular frequencies and
durations in the usage of fixations in unconstrained interactions
could help us understand patterns in attention to targets.

As mentioned previously, there is ample evidence that
attention to social sounds (such as calling a child by her name)
is severely impaired in children with ASD ([53], [52], [51],
[11]). This impairment could exacerbate the difficulties that
people with PDD face when simultaneously processing visual
and auditory stimuli ([22], [23], [24], [25], [26]). Therefore
we set forth to investigate whether children with PDD tend
to gaze at an adult’s face differently when the adult talks
to them, from when the adult makes a facial expression of
emotion. Because the social difficulties of children with ASD
in everyday settings are more pronounced than in experimental
tasks ([37], [66]), we conducted our study at the children’s
homes and schools while they were playing familiar games
with a familiar adult.

In this study, we contrasted the gaze patterns directed to
an adult’s face and toys in a group of 13 children with PDD,
between 3 and 11-years old, and 13 TD children, between
2 and 6-years old. We report on a global analysis of the
proportion, frequency and duration of gaze directed to faces
during a complete play interaction; we test the hypothesis that
children with PDD will look at the adult’s face less than TD
dependent on gender.

In our previous work, Magrelli et al. [67], we investigated,
in children with ASD, the patterns in attention orienting to
social stimuli during dyadic social interactions taking place
in real-life settings. Specifically, we focused on analyzing the
temporal gaze patterns directed to the adult’s face immediately
before and after the onset of the presented social stimuli
that differ in complexity: social cues produced by facial
expressions of emotion and those produced during speech.
We recorded the children’s gazes by using the WearCam eye-
tracking device and we report on a spatio-temporal (frame
by frame) analysis of the gaze prior and after onset of the
two social events. The study involved a group of children
with ASD from 2 to 11-years old (n = 14) and a group of
typically developing (TD) children (n = 17) between 3 and
6-years old. We found that both groups oriented overtly to
facial expressions, though children with ASD did so to a lesser
extent. Moreover, children with ASD shifted their attention to
speaking faces proportionally less and more slowly than TD
children. These findings suggest that people with ASD show
reduced overt shifting of attention toward social stimuli. These
results also support the hypothesis that individuals affected by
ASD have difficulties processing complex social sounds and
detecting intermodal correspondence between facial and vocal
information.

II. Method

A. Participants

Study of the group: A group of children with PDD (n=13)
was compared with a group of TD children (n=13). The
children with PDD were diagnosed using the DSM-IV-TR
criteria [1] by medical doctors and clinical psychologists at
the Geneva University Hospital. They were scored with the
CARS test ([68]) as presenting mild to severe autism (Mean
score: 39.62 ± 6.60, range: 30-49). Their mean chronological
age was 6.17 ± 2.40 years, (range: 3.08 - 10.6). The mean
chronological age of the TD children was 3.68 ± 1.29 years
(range: 2.33-6.08), see Table I for individual data. The two
groups of children were matched on gender.

The participants with PDD were recruited through the
Geneva University Hospital and through the Autisme Suisse
Romande Association. Participants in the control group were
recruited at the daycare facilities of EPFL. All parents gave
their written informed consent, including permission to use
video-recordings for scientific publications. The protocol and
form of consent were approved by the ethics committee of the
University Hospital of Geneva.
Only one child with PDD was uncooperative. Hence, another child was recruited to complete the population. Note that, if a child was uncooperative during the first visit, we gave her a second chance, and we scheduled a second visit within two weeks after the first. Four children with PDD were uncooperative during the first visit, but subsequently cooperated during the second visit. No subject had to be removed because of technical problems. None of the control subjects refused to wear the Wearcam.

### B. Apparatus: The WearCam System

The WearCam is a wearable head-mounted eye-tracking system, especially designed to provide a light and non-invasive device that would be suitable for young children, see Figure 1. An elastic band with velcro straps allows for the WearCam to be tightly secured to the child’s head. The system is lightweight and easy to wear and does not obstruct the child’s field of view.

The WearCam is composed of two Sony Super HAD CCD cameras mounted with a mirror, see Figure 1 (field of view: 96° horizontal by 96° vertical). The first camera, located on top, is aligned with the direction of the child’s head and thus gives a view of the “interaction zone”, i.e., the area the child sees when her eyes are aligned with her head, such as when looking at people and objects from afar. The second camera, located below the first, forms an angle of 45° with the direction of the head and gives a view of the “lower zone”, i.e., the area the child sees when looking down. A mirror mounted on the lower camera captures the reflection of the child’s eyes that appears at the top of the image of the bottom camera, see Figure 2. Recordings from the Wearcam encompass two images showing, on the bottom of the image, the eyes of the child and, on top of the image, the capture of the field of view (the images of the scene from the top and bottom cameras are merged to create a single image). The reflection of the child’s eyes in the camera is used during the post-hoc analysis of the video data, to track the child’s gaze during the calibration procedure.

The eye-tracking accuracy of the Wearcam was assessed in a separate study by Noris et al. [63] on 10 typically developing children (6 girls, 4 boys, mean chronological age 2.4 ± 0.4 years). Angular precision of 2.42° ± 0.89 was revealed. A comparison with 12 other eye-tracking systems, including the Tobii and the head-mounted systems from ISCAAN and ASL ([63]) showed that optimal accuracy can be obtained with a minimum of 70 calibration points. Two kinds of calibration are available with the Wearcam: on site calibration [69] and off-line calibration [63]. In our experiments, we performed an off-line calibration. Calibration was done using the first 2-3 minutes of recordings prior to the beginning of the study, when the child was left to play with a little toy to divert her attention. A trained rater visualized the video of the field of view and of the eyes by using a custom-made software. The experimenter used all identifiable instances in which the direction of the child’s gaze was unambiguous (e.g., when the child reached toward an object and the eyes shifted toward it), and she placed a calibration point at the corresponding position in the image. This is possible because the eyes of the child are constantly visible in the recorded mirror. For each recording session and for each child, a set of 70 to 100 data-points were gathered by two trained raters with an inter-rater correlation of 0.9, see Table II. Further details on the calibration process are explained in Noris et al. ([63]).

After, a non-linear mapping is computed between the two-dimensional coordinates on the image of the field of view and on the image of the eyes, returned by the Wearcam at the same instant. Mapping the calibration points to the complete image of the visual field returned by the cameras is done through non-linear regression by using support vector regression (SVR) ([70]). An SVR was trained using a RBF kernel with \( \sigma = 0.01 \), \( C \)-intensive loss function with \( C = 100 \), and soft margin constraint \( C = 1000 \). The SVR was performed separately on the horizontal and vertical dimensions of the image. Note that the SVR mapping does not rely solely on geometrical elements, such as position of the iris and pupil; it also exploits other features such as the shape of eyelashes and shading on the eyelids. This allows the system to extract information on gaze direction, even when the child is looking downwards and the iris is not completely visible. To adapt the SVR model to the fact that the eyes appear differently in the image, we gather a few additional calibration points by using the part of the video recordings that follows the displacement of the hat. Specifically, every time a headset slippage occurs during an experiment, we let the child play with a game for 2-3 minutes and collected additional calibration points. This procedure was identical to the one conducted during the post-hoc calibration phase that we conducted at the beginning of the experiment.

Note that currently some available wearable eye-tracking systems are sufficiently light to be used by children but they require the child to undergo a calibration procedure prior to the study. This assumes a cooperative user and can be difficult to request from children with PDD. The Wearcam eye-tracking technology, however, does not require a pre-
We designed a simple behavioural coding system for studying the gaze behaviors adopted by the children during the naturalistic interactions and for examining whether they were affected by the complexity of the social stimuli. Double-blinded raters, blind to the diagnosis and to the aim of the study, were asked to label the video sequences in which the adult interacting with the child was either speaking or making facial expressions with the explicit intent of communicating with the child, see Table II. Therefore, for an event to be labeled respectively as speech or a facial expression, the adult had to look directly in the child’s eyes and, either speak or show a feeling (e.g., surprise or joy). Note that we explicitly required that the events labeled as facial expression should not temporally overlap, nor immediately follow or precede the events that were labeled as speech. The inter-rater reliability for the labeling of events was measured this time on 27% of the data using intra-class correlation. Results are reported in Table II.

After the rating process, we isolated four events of interest that were analyzed, corresponding to four different experimental conditions:

1) **Face** condition: the subset of video frames containing the adult’s face in the field of view, regardless of the kind of interaction taking place and the presence of objects in the scene.

2) **Facial expression** condition: the subset of video frames in which the child held the adult’s face in the field of view, while the adult was expressing emotions with facial expressions and no objects were in the field of view. Three facial expressions were considered: smile, surprise and interrogative. Note that, in all the examined cases, these expressions could be associated with positive emotions.

3) **Speaking** condition: the subset of video frames in which the adult’s face was visible, while the adult was speaking and no objects were in the field of view.

4) **Object** condition: the subset of video frames in which the object used in the experiment was visible in the field of view and the adult’s face was not visible in the field of view.
In both the speaking and facial expression conditions, we encompassed only the scenes where none of the experimental toys were visible in the field of view of the camera. The rationale behind this selection is to avoid bias on the measure of the gaze: children could have been focusing on objects present in their field of view rather than the adult’s face. By avoiding distractors, we can precisely describe behaviors in response to social cues. Similarly, in the object condition only the frames in which the adult’s face was not visible in the child’s field of view are taken into account, in order to avoid biases on the measure of the gaze. In addition, to investigate the general tendency of children towards faces in the presence of distractors and without discriminating the intensity of the social cues, we analyze the gaze behaviors over the whole course of the play interaction when adult’s face is present (face condition).

We did not require raters to make a distinction between specific kinds of facial expressions. First, we observed that during the experiments, the adults displayed only facial expressions with positive rewards, as would be naturally expected from a play session. Although an extensive analysis of gaze patterns towards specific facial expressions and utterances could be of great interest, we compared only facial cues with or without auditory stimuli. An in-depth analysis of attentional responses to different facial expressions or speech acts would require building a specific protocol that balances the frequencies of each kind of event across the experiments. This would constrain the interactions between the children and the adults, thus reducing the spontaneity of the responses to socially demanding situations in which we are interested.

E. Data Processing in the Child’s Field of View

We report on a comparative analysis of the gaze patterns between the PDD group and the control group, see Table III.

The motion of the child’s eyes was tracked and the locus of her gaze was reconstructed in the complete image returned by the WearCam. Knowing the location of an object of interest, returned by the tracker, enables us to determine whether the object of interest is looked by the child as an effect of how far the corresponding area is from the camera (by using the size of its bounding box returned by the object detector and computing a homothetic transformation). For instance, a face is considered to be looked by the child, if 2/3rds of the width of its bounding box is contained within the circular sector associated with the foveal vision, i.e., if it is between 0° and 3° of eccentricity.

F. Data Analysis

We adopt the definition of Gaze Episode described by Noris et al. [72]. A Gaze Episode is defined as the span of time between the instant (image frame) the gaze moved on the bounding box corresponding to an object of interest and the instant it left the bounding box. In our analysis, we considered only gaze episodes that are at least 120 ms long (equivalent to 3 frames) to avoid counting short fixations and movements that crossed the object of interest but did not linger there. Gaze episodes were used to avoid the drawbacks related to the explicit computation of fixations (see [73] for a thorough discussion of this issue).

Our analysis of the data is based on a statistical comparison of five (Xi, i = 1, . . . , 5) within-subject numerical gaze variables that measure the proportion of time, region, frequency and duration of Gaze Episodes that the TD children and children with PDD spent looking at the areas of interest in the four conditions face, facial expression, speaking and object, described in section D. Hence, a Gaze Episode toward human faces, for instance, is labelled “in Fovea”, and is taken into account in the statistical analysis, when a face is present inside the fovea of the child (i. e., across the macula, a 3° region around the gaze direction).

Specifically, we compute the following four outcomes:

1) EpPerc Face: proportion of time the child gazes at the adult’s face, normalized by the time during which the adult’s face appears in the field of view, in the face condition (X1,Face).

2) EpPerc Speaking: proportion of time the child gazes at the adult’s face, normalized by the time during which the adult’s face appeared in the field of view, in the speaking condition (X1,Speak).

3) EpPerc Facial Expression: proportion of time the child gazes at the adult’s face, normalized by the time during which the adult’s face appears in the field of view, in the facial expression condition (X1,FacExp).

4) EpPerc Object: proportion of time the child gazes at the toy, normalized by the time during which object of interaction appears in the field of view, in the object condition (X1,Obj).

The frequency of Gaze Episodes is computed as the number of occurrences of Gaze Episodes on the area of interest per minute and is recorded in Hz. To ensure that the measurement of frequency is computed on an equal scale for all recordings and for all children, the frequency is computed on segments of 40 seconds (in the results section, we report on the mean value of the frequency across the segments). We also compute the relative frequency and duration. Specifically, we computed the following outcomes:

1) EpFreq Face: frequency of Gaze Episodes, in the face condition X2,Face ;

2) EpFreq Speaking: frequency of Gaze Episodes, in the speaking condition, X2,Speak ;

3) EpFreq Facial Expression: frequency of Gaze Episodes, in the facial expression condition,
$X_{2,FacExpr}$;

4) **EpFreq Object**: frequency of Gaze Episodes, in the object condition, $X_{2,Obj}$;

5) **EpDur Face**: duration of Gaze Episodes, in the face condition, $X_{3,Face}$;

6) **EpDur Speaking**: duration of Gaze Episodes, in the speaking condition, $X_{3,Speak}$;

7) **EpDur Facial Expression**: duration of Gaze Episodes, in the facial expression condition $X_{3,FacExpr}$;

8) **EpDur Object**: duration of Gaze Episodes, in the object condition ($X_{3,Obj}$);

In addition, we computed two other measures that represent the general dispersion of faces in the broad field of view for the four conditions:

1) **InFovPerc**: Proportion of time a face appears in the broad field of view, in the different conditions ($X_{4,Face}, X_{4,Speak}, X_{4,FacExpr}, X_{4,Obj}$);

2) **InMiddle**: Proportion of time a face appears in the middle of the field of view, i.e., when the center of the bounding box of the face is located within a circular sector of radius $3^\circ$ of eccentricity and centered in the middle of the image. We computed this measure in all the conditions: ($X_{5,Face}, X_{5,Speak}, X_{5,FacExpr}, X_{5,Obj}$);

The influence of the two categorical factors, diagnosis and task, and of the ordinal factor, chronological age, were measured through a mixed design Analysis of Variance (ANOVA) with three between-subject factors: diagnosis (PDD vs. TD), task (play-doh vs. bubble) and chronological age (continuous value). For all statistical tests, we highlight the effects deemed significant, i.e., with a $p$-value smaller than 0.05 and report eta squared as a measure of effect size ([74]).

III. RESULTS

We compare the proportion ($X_1$) of time the child’s eyes look at an area of interest, normalized by the time during which the area of interest appears in the child’s field of view (In FoV), between the PDD and TD groups. We also report comparisons between the groups also on the Gaze Episode frequency, ($X_2$), duration, ($X_3$), and on the general dispersion of faces in the broad field of view: In Field-of-View Percentage ($X_4$) and In Middle ($X_5$).

In this study, we perform the comparisons for the four different conditions: (a) **Face** - over the course of the whole interaction, (b) **facial expression** - while the adult was making facial expressions and no object was present in the child’s visual field of view, (c) **speaking** - while the adult was speaking to the child and no object was present in the child’s visual field of view and (d) **object** - while the objects of interaction were present in the child’s field of view and faces were absent.

A. Study of the Group

The current study provides a comparison between two samples of 13 children with PDD and 13 TD children. For each child, we observe the child playing the bubbles and the play-doh.

### Table III

<table>
<thead>
<tr>
<th>FACE CONDITION</th>
<th>( \text{TD} )</th>
<th>( \text{PDD} )</th>
<th>ANOVA(Diagnosis)</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{5,FoV}$</td>
<td>63.64% ± 24.24</td>
<td>52.02% ± 25.14</td>
<td>p = 0.070 F(1,57) = 3.45</td>
<td>.054</td>
</tr>
<tr>
<td>$X_{6,InMiddle}$</td>
<td>9.21% ± 13.76</td>
<td>8.44% ± 8.52</td>
<td>p = 0.789 F(1,57) = 0.07</td>
<td>.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACIAL EXPRESSION CONDITION</th>
<th>( \text{TD} )</th>
<th>( \text{PDD} )</th>
<th>ANOVA(Diagnosis)</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{1,EpFreq}$</td>
<td>7.64% ± 3.45</td>
<td>4.51% ± 2.67</td>
<td>p = 0.00 F(1,57) = 18.45</td>
<td>.211**</td>
</tr>
<tr>
<td>$X_{3,EpFreq}$</td>
<td>8.71% ± 4.73</td>
<td>4.96% ± 3.81</td>
<td>p = 0.00 F(1,57) = 17.66</td>
<td>.188**</td>
</tr>
<tr>
<td>$X_{4,EpDur}$</td>
<td>.34 ± 0.06</td>
<td>.31 ± 0.08</td>
<td>p = 0.198 F(1,57) = 1.71</td>
<td>.032</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SPEAKING CONDITION</th>
<th>( \text{TD} )</th>
<th>( \text{PDD} )</th>
<th>ANOVA(Diagnosis)</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{1,EpFreq}$</td>
<td>2.55% ± 1.85</td>
<td>4.07% ± 6.20</td>
<td>p = 0.273 F(1,41) = 1.25</td>
<td>.028</td>
</tr>
<tr>
<td>$X_{6,InMiddle}$</td>
<td>.86% ± 1.31</td>
<td>1.08% ± 2.05</td>
<td>p = 0.667 F(1,41) = 0.19</td>
<td>.004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACIAL EXPRESSION CONDITION VS SPEAKING CONDITION</th>
<th>( \text{PDD} )</th>
<th>( \text{FE} )</th>
<th>( \text{S} )</th>
<th>ANOVA(Diagnosis)</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{1,EpFreq}$</td>
<td>8.23% ± 5.53</td>
<td>5.08% ± 2.53</td>
<td>p = 0.012 F(1,41) = 7.21</td>
<td>.123**</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OBJECT CONDITION</th>
<th>( \text{TD} )</th>
<th>( \text{PDD} )</th>
<th>ANOVA(Diagnosis)</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{5,FoV}$</td>
<td>35.54% ± 24.04</td>
<td>31.62% ± 20.98</td>
<td>p = 0.388 F(1,52) = 0.76</td>
<td>.007</td>
</tr>
<tr>
<td>$X_{6,InMiddle}$</td>
<td>14.68% ± 14.54</td>
<td>10.14% ± 9.03</td>
<td>p = 0.078 F(1,52) = 3.24</td>
<td>.032</td>
</tr>
</tbody>
</table>

Analysis of How the Gaze Span the Field of View: A three-way mixed ANOVA, conducted independently for each gaze variables with diagnosis, task and chronological age as independent variables, yields a significant effect for diagnosis in the face condition, specifically for the proportion of time of looking the adult’s face, see Table III. Moreover, in the face condition the frequency of Gaze Episodes ($X_{2,Face}$) differs significantly for the two groups. Recall, the face condition refers to all the experimental sessions, regardless of the kind of interaction and the presence of toys in the scene. In particular, we observe that TD children looked directly at faces proportionally more time than children with PDD, ($X_{1,Face}$), TD: 7.64% ± 3.45, PDD: 4.51% ± 2.67, p < 0.001.
Moreover, children with PDD looked at the adult’s face less frequently than the TD ones did, \( (X_{2,Face} [40,96]) \), (TD: 8.71% ± 4.73, PDD: 4.96% ± 3.81, p<0.001, \( F=(1,57)=17.76) \).

We observe a large-effect size for the variables EpPerc (Gaze Episode percentage, \( X_{1,Face} \)), the relative EpFreq (Gaze Episode frequency \( X_{2,Face} \)). Conversely, in facial expression and speaking conditions, we find a significant difference between the two groups, only for the proportion of Gaze Episodess on the adult’s face. Specifically, in the comparison between the proportions of times the children looked directly at the adult’s face in the speaking condition (\( X_{1,Speak} \)), there is a large-effect size. (TD children looked on speaking faces significantly more than those with PDD did (TD: 10.62% ± 10.7, PDD: 5.08% ± 2.53, F(1,46)=7.17, p: 0.001)). Similar proportions of Gaze Episodess (\( X_{1,FaceExpr} \)) can be observed in the facial expression condition with middle-effect size (PDD: 11.96% ± 7.11; TD: 8.23% ± 5.53, F(1,41)=5.04, p=0.032).

The PDD group does not differ from TD group in the frequency with which they look at faces, in the two conditions speaking and facial expression. A significant difference in the gaze patterns in children with PDD between the facial expression condition and speaking condition is present for the proportion of Gaze Episodess to the adult’s face, with large-effect size (\( X_{1,FaceExpr} \): 8.23%±5.53, \( X_{1,Speak} \): 5.08%±2.53, p:0.011, F(1,46)=7.21), as shown at the bottom of Table III.

It is important to note, we do not observe main or interaction effects on the chronological age and task for any of the gaze measurements.

Table III compares values obtained for each gaze variable for the PDD and TD groups. TD children and children with PDD did not differ in the percentage of faces (\( X_{4,Face} \), \( X_{4,FaceExpr} \), \( X_{4,Speak} \) appearing in their broad field of view in all the three conditions (>50% in face, ~4% in facial expression, ~4% in speaking). As there was no difference in the percentage of time the adult addressed the child verbally or through an expression of emotion. This supports the assumption that the level of interactivity of the adult was the same for both groups.

In the object condition, we did not observe any significant main or interaction effect of the three factors for the gaze strategies between the two groups. Children with PDD and TD children did not differ in the proportion of Gaze Episodess falling on the toy (\( X_{1,Obj} \)) used during the interaction. Similarly, the frequency (\( X_{2,Obj} \)) and duration of Gaze Episodess (\( X_{3,Obj} \)) did not differ significantly for the two groups.

The Effect of the Severity of Pervasive Developmental Disorders: For the face condition, our results indicate that the children with PDD looked less directly at the adult’s face than the control children, and when they did so, they looked less frequently though for similar durations. To test whether this paucity of gaze directed to the face is aggravated as a function of the severity of autism spectrum disorders, we computed a measure of the correlation between the CARS score and the corresponding values obtained for the gaze measurements, for all children affected by PDD in the three conditions: face, facial expression and speaking. Correlation coefficients were not statistically significant for EpPerc (Gaze Episode percentage), EpFreq (Gaze Episode frequency) and EpDur (Gaze Episode duration).

The Effect of Task: As stated previously, a three-way ANOVA with Task as a dependent factor yielded no main effect, nor interaction effects on any of our gaze variables. The absence of an effect of Task on variables In FovPerc and In Middle, in the face condition, means that the proportion of faces appearing in the global field of view and in the middle of the visual field was similar for both tasks. This is evidence that participants interacted visually in a similar manner in both tasks.

IV. DISCUSSION

Our results confirm the literature that reports impaired overt attention to social stimuli in autism spectrum disorders (75], [53], [76], [59], [9], [32]) in naturalistic settings. Indeed, we find that, in general, children with PDD looked proportionally less time at faces and less frequently than TD children did. This happened during the whole set of experiments, regardless of the nature of the social interaction and whether or not objects were present in the scene (i. e., in the face condition). As explained in the Introduction, attention impairments in face and facial expression processing have been found to be contradictory in individuals with ASD where experiments are conducted in highly controlled settings ([42], [8], [14]). In this sense, the face condition, being completely unconstrained, offers us the opportunity to analyze the spontaneous selection of social cues in complex ambiguous situations where individuals with ASD show severe social attention deficits [14]. As Kiln et al. [37] state ”despite three decades of experimentation, the most obvious indication of the profound social disability witnessed in autism is still the spontaneous presentation of affected individuals in unconstrained social interaction”. Eye-tracking technology allows us to quantify attentional deficits and therefore it has the potential to be used for the diagnostic assessment of ASD [77]. We highlight one of the clearest advantages of the WearCam eye-tracker over others: Its ability to capture the dynamics of gaze patterns with precision, combined with its unobtrusiveness could be helpful in testing controversial hypotheses on autism spectrum disorders in complex scenarios [14]. This capacity would provide researchers with a better understanding of the nature of social attention abnormalities in PDD.

In our study, the two groups of children do not differ statistically in the duration and frequency of episodes of looking at faces, when only social cues (speech, facial expression) are presented and distracting objects are not present (i. e., both in the facial expression condition and in the speaking condition). However, under all these conditions, children with PDD looked directly at the adults proportionally less time than TD children did. Moreover, the children affected by PDD looked at faces in the speaking condition proportionally less than in the facial expression condition. This fact could indicate a tendency to avoid faces during speech and it could be explained by the deficits in speech auditory processing, highlighted in numerous studies ([9], [59], [60], [61]). In addition, this result seems to be in line with the retrospective
studies that report a lack of social orienting in children with autism ([51], [52], [53]). It is important to underline that, the co-occurrence of two social stimuli: visual and auditory, require multisensory processing of the social cues. Impairment in processing of such multimodal social information could be more important and more visible than impairment in processing unimodal information (see Iarocci et al. [78], for a review). For this reason, the measurement of impairment in multimodal processing skills of social stimuli could help in early detection of PDD. In fact, these skills appear during the first 6 months of life and accompany well-documented stages of development in the TD population ([79], [80], [81]). To our knowledge, only a few studies quantitatively compare attention to stimuli of different social complexity in autism spectrum disorders in real-life settings (e.g., by analyzing retrospective home videos). Note that our results agree with the recent eye-tracking studies conducted in semi-naturalistic settings by Chawarska et al. [54] and Shic et al. [55]. Both studies reported that children with ASD looked proportionally less time at an actor’s face, when the actor was speaking, than TD children and children with developmental delays.

In our study, decreased looking time at social cues expressed by faces in children with PDD did not result in an increased looking time at objects. This finding suggests that the lack of looking time towards social stimuli might not be due to the fact that non-social objects are more salient or more interesting for children with PDD. Our study complements the literature dedicated to overt attention by providing information on how people with PDD attended to social stimuli in naturalistic interactions with familiar adults. Our study did not control developmental age. This is a limitation of the present study. In fact, we cannot assess whether the gaze strategies for looking at the adult’s face become more or less relevant with developmental age.

This study replicates our earlier findings [67] related to the analysis of spontaneous shifting of attention in response to the occurrence of two social stimuli: facial expressions and speech. Our previous study involved a group of 14 children with ASD from 2 to 11-years old and a group of 17 TD children between 3 and 6-years old. It was conducted using live recordings during naturalistic interactions similarly to those documented in this study. Although the two groups oriented overtly to facial expressions, children with ASD did so to a lesser extent. The children with ASD significantly differed from TD children in their way of shifting overtly their attention to speaking faces. When children with ASD orient to facial expressions, their reaction times and first fixation lengths were similar to those presented by TD children. To speaking faces, however, children with ASD oriented more slowly than TD children. Whereas this study quantifies the amount of overt attention towards facial expressions and speech, our previous work focused on identifying peculiarities in overt shifting of attention towards the same kind of social stimuli. A similar investigation was previously conducted by Sasson et al. [82] and focused on investigating social orienting to facial expressions. Sasson et al. contrasted the gaze behaviors of adults with autism, with those of adults with schizophrenia and with those of typically developed peers. The subjects observed a series of complex social scenes where facial expressions of emotion were either included or digitally erased. They report that both the subjects with autism and with schizophrenia look directly at faces to a lesser extent than the TD individuals did. Moreover, subjects with autism were found to orient at the same velocity when emotionally meaningful information were both presented and erased. However, the authors did not report any difference in the latency to look at the adult’s facial expressions, between the TD group and the subjects with autism.

Finally, even though, an extensive analysis of gaze patterns, towards a particular subset of facial expressions and speech acts, can be of great interest in naturalistic settings; this experiment was not designed specifically to test for these cases and the occurrence of each facial expression was too small to provide for enough statistics to conduct such a detailed analysis.

V. CONCLUSION

In our experiments, we compared gaze exploration strategies during playful social interaction. The study involved a group of 3-to-11-years old children with pervasive developmental disorders (PDD) compared to a group of typically developing (TD) children between 2 and 6-years old. We contrasted the gaze behavior of the two groups of children towards faces, regardless of the social interaction offered by the adult and the presence of objects. The children with PDD looked directly at the adult’s face, proportionally less, and less frequently than TD children did.

We further narrowed our investigation to contrast gaze strategies in the presence of two types of social stimuli, specifically when the adult verbally addressed the child or through a facial expression of emotion. In this case, to analyze the effect of the social stimuli alone, we excluded from the analysis the segments in which objects were in the child’s field of view. This last investigation confirms the presence of atypical gaze strategies in children with PDD, independently of the gender of the child. Similarly to the previous condition, children with PDD directly gazed proportionally less time although with the same frequency at others than typical children did.

Our study is limited in the number of subjects and in the fact that we did not measure developmental age for the sample with PDD. Whereas the results are encouraging and of importance, this represents only a preliminary study and it is necessary to test whether the gaze factors that discriminate children with PDD from controls in this investigation are specific to the PDD perceptual peculiarities or can be found in other clinical groups. Extensive investigations are required to better assess the characteristic abilities and processing behind attention strategies adopted in natural settings in both TD people and individuals with PDD.

In general, the literature on studies conducted in natural settings is scarce when compared to the research effort dedicated to the study of attention in autism spectrum disorders in laboratory, artificial settings. However, research in ASD has recently stressed the importance of investigating the nature of social attention in real life settings where manifestations of
imperfections in ASD are more manifested. The precision in quantifying social attention via a head-mounted eye-tracking system can help us to understand the nature of attentional deficits in ASD. In addition, eye-tracking technologies can be used to assist the diagnostic process of ASD. The two crucial advantages of the eye-tracking technology we used, the WearCam device, are its unobtrusiveness and portability. Due to these key features of the WearCam technology, we were able to capture the dynamics of gaze behavior during complex social interactions, such as play sessions with children. Future studies could focus on using the WearCam system to test controversial hypotheses in ASD, such as those related to peculiarities in covert attention and covert shifting of attention in unconstrained settings.

ACKNOWLEDGMENTS


REFERENCES


