information

MDPI

Article

Requirements for Robotic Interpretation of Social Signals "in the Wild": Insights from Diagnostic Criteria of Autism Spectrum Disorder

Madeleine E. Bartlett [1,*], Cristina Costescu [2], Paul Baxter [3] and Serge Thill [4,5]

1 Centre for Robotics and Neural Systems, University of Plymouth, Plymouth PL4 8AA, UK
2 Department of Clinical Psychology and Psychotherapy, Babes-Bolyai University, Cluj-Napoca 400000, Romania; christina.costescu@gmail.com
3 Lincoln Centre for Autonomous Systems, School of Computer Science, University of Lincoln, Lincoln LN6 7TS, UK; pbaxter@lincoln.ac.uk
4 Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, 6525 XZ Nijmegen, The Netherlands; s.thill@donders.ru.nl
5 Interaction Lab, School of Informatics, University of Skövde, 541 28 Skövde, Sweden
* Correspondence: madeleine.bartlett@plymouth.ac.uk

Received: 12 December 2019; Accepted: 30 January 2020; Published: 1 February 2020

Abstract: The last few decades have seen widespread advances in technological means to characterise observable aspects of human behaviour such as gaze or posture. Among others, these developments have also led to significant advances in social robotics. At the same time, however, social robots are still largely evaluated in idealised or laboratory conditions, and it remains unclear whether the technological progress is sufficient to let such robots move "into the wild". In this paper, we characterise the problems that a social robot in the real world may face, and review the technological state of the art in terms of addressing these. We do this by considering what it would entail to automate the diagnosis of Autism Spectrum Disorder (ASD). Just as for social robotics, ASD diagnosis fundamentally requires the ability to characterise human behaviour from observable aspects. However, therapists provide clear criteria regarding what to look for. As such, ASD diagnosis is a situation that is both relevant to real-world social robotics and comes with clear metrics. Overall, we demonstrate that even with relatively clear therapist-provided criteria and current technological progress, the need to interpret *covert* behaviour cannot yet be fully addressed. Our discussions have clear implications for ASD diagnosis, but also for social robotics more generally. For ASD diagnosis, we provide a classification of criteria based on whether or not they depend on covert information and highlight present-day possibilities for supporting therapists in diagnosis through technological means. For social robotics, we highlight the fundamental role of covert behaviour, show that the current state-of-the-art is unable to characterise this, and emphasise that future research should tackle this explicitly in realistic settings.

Keywords: autism spectrum disorder; diagnosis; technology; behaviour

1. Introduction

Having robots engage socially with humans is a desirable goal for social robotics. It lowers the barrier to entry into interactions, as it allows the humans to engage and interact with the robot in a way similar to how they would interact with another human. This would remove the need for any specialist robotics knowledge or training for the human users, and thus substantially expands the application domains for social robots beyond the current largely restricted environments in which they are currently used. However, there remain a range of fundamental challenges to being able to

achieve this. Principal among these is that in order to behave appropriately, it is necessary for the robot to *understand* what its human interaction partner is doing (and indeed what they may do). Apart from current limitations in sensory detection technologies (which are improving), the problem remains that essentially the robot observer can only have information about observable (overt) behaviour, but has no access to the mental states (or covert aspects of behaviour) that led to these overt behaviours – this requires further inference. This fundamental challenge for social robotics is the topic of this contribution: we characterise the current state of the art with respect to this problem, synthesising advances across a range of technology disciplines, and highlighting where further technological advances can be most usefully made.

1.1. Recognising Human Internal States from Observable Kinematics in Social Robotics

The ability to infer the mental states of other agents is a fundamental component of social interaction. In humans, this ability is called "Theory of Mind". The exact mechanisms underlying it remain unclear; some hypotheses center around an ability to create folk-psychological models of other minds while others suggest that internal simulation mechanisms normally used to control one's own behaviour can be used to understand and predict the behaviours of others from observation [1,2]. In robotics, the latter, along with its connections to mirror neurons, has long inspired, for example, forms of imitation learning and action understanding that rely on the robot's own forward and inverse kinematic models [3–6].

That said, merely predicting the outcome of actions is not the same as understanding internal mental states from observable kinematics. The latter is seen as a pre-requisite for truly social robotics, yet remains a challenge [7]. While we will give a brief overview of relevant work in the sections below, much current work in social robotics does not address this directly but focuses on, among others, characterising end user requirements in specific applications [8] or studying the degree to which phenomena known from social sciences are applicable to human-robot interactions [9]. It is noteworthy that relatively little is actually required of the robots themselves in such studies, and a Wizard-of-Oz control paradigm is sufficient. Applications of social robots that do require the robot to possess at least some autonomous behaviour exist, for example in education [10] or robot-assisted therapy for disorders such as Autism Spectrum Disorder [11–13], but these are still relatively narrow domains within social robotics.

Overall, there is relatively little research that directly investigates the degree to which the state of the art currently allows social robots in the more general sense. At the same time, this is a timely question since, as we will discuss in this paper, technological progress in recent years does allow for relatively comprehensive observation of human agents in the environment and, together with advances in data analysis (for example, using deep networks) is at a point where it might be feasible to advance in this direction as well.

In this paper we evaluate this technological progress and the degree to which it fulfils the needs of social robots that would exist "in the wild", and not constrained to narrow domains. To perform such an evaluation requires a scenario that captures the essential requirements for social robotics. Here, we focus on the automation of the *diagnosis* of Autism Spectrum Disorder (ASD) for this purpose. We will detail this problem domain further below, but it is important to note that it is distinct from using robots in ASD *therapy*: indeed, diagnosis, in principle, does not even require a robot. On the other hand, diagnosing ASD does require the ability to observe social interactions and infer underlying mental states, which is the core requirement for social robots that we are interested in here. It is also a domain for which clear protocols, assessment criteria and so on exist. For our purposes, this is a crucial advantage over other social contexts because it provides us with the ability to evaluate the degree to which technology can meet these criteria. It is also worth noting that the automation of ASD diagnosis is in itself a relevant research topic; not to replace the clinical therapists involved, but to support them: as we will see below, the process is rather intensive but opportunities for alleviating the burden exist.

In the remainder of this introduction, we first describe ASD and diagnostic criteria. We then break these down into different categories, based on whether they focus just on the behaviour of the child or on the interaction itself, and whether they concern the assessment of overt or covert information. We then discuss the degree to which technological means can fulfil these requirements.

1.2. Diagnosing Autism Spectrum Disorder

ASD is characterised by the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) [14] using two categories of behaviour: social communication difficulties and restricted or repetitive behaviour patterns. Since the identification of ASD [15,16], the literature has examined potential causes, intervention techniques and approaches to diagnosis. These investigations have revealed ASD to be a complex developmental disorder with high levels of heterogeneity within the clinical population in terms of symptom presentation and severity [17]. Furthermore, there are no biologically based tests for ASD [18]. As such, the diagnosis of ASD remains a very difficult task, relying on the interpretation of current and retrospective observations of an individual's behaviour, and of developmental aspects, by different specialists including psychologists, psychiatrists and speech therapists [19,20]. These observational judgements are then quantified according to standard protocols such as the Diagnosis Interview Revised (ADI-R) [21], the Childhood Autism Rating Scale (CARS) [22], and the Autism Diagnostic Observation Schedule Generic (ADOS-G) [23].

Despite the efforts made thus far to improve and standardise the diagnostic process (via the tools listed above), the variable nature of ASD and the emergence of symptoms in early childhood [24] amid ongoing developmental changes does cause difficulties for its identification and diagnosis [18]. While standardisation of the diagnostic process via tools such as those above has been effective in aiding clinicians in this task [20,25], there is room for improvement. In particular, surveys asking parents about the process of getting an ASD diagnosis for their child found that even though parents first seek a diagnosis when their child is aged 3.9 years (on average), a final diagnosis is not received until the child is 7.5 years. Consequently, one way in which the diagnostic process could be improved would be to reduce the time taken from when parents first seek a diagnosis to when a final diagnosis is received [26].

One way to address this would be to provide protocols which are easier to implement, and able to produce useful information without over-reliance on human expertise and thereby provide General (GPs) and other practitioners with the means to make more informed decisions about when to refer a patient for expert diagnosis. It is important to note that we do not propose to replace the assessments carried out by expert clinicians, but rather to make the process of accessing these assessments easier, cheaper and more efficient. We propose that technologies able to provide useful information about an individual's diagnostic status could contribute to achieving this goal.

Technical advances have long inspired research into how technologies can be applied to diagnostic scenarios, a method referred to in the medical field as Computed-Aided Diagnosis (CAD) [27]. These applications have various motivations including improving the objectivity of decision-making or measurement [28] and incorporating information into the diagnostic process that is more readily detected, measured or used by computers than humans alone [27]. While such techniques were applied to physiological maladies, the advent of technologies and methods for measuring human behaviours, e.g., via machine-perception-guided technologies, has created opportunities for augmenting the diagnosis of behavioural and psychological disorders such as ASD.

2. Observable Behavioural Cues

The first step in augmenting the diagnosis of ASD with technology is to identify whether there are any diagnostic markers that existing technologies can measure and quantify in a meaningful way. To do this we must first identify symptoms that have been sufficiently operationalised to provide objective definitions. Arguably, the DSM and existing diagnostic tools provide such definitions. Support for this claim comes from tests of the reliability and objectivity of these definitions via Inter-Rater Agreement

(IRA). Several studies have looked at IRA between clinicians on the items included in diagnostic tools. While evidence shows that IRA for observational judgements is typically low [29], recent studies examining IRA for clinicians' diagnostic evaluations using the ADI-R [30] and ADOS [31] tools (whose symptom definitions are based on those provided by the DSM) have demonstrated high levels of agreement for each behavioural marker outlined by each tool. These findings demonstrate that the DSM has successfully operationalised the diagnostic characteristics of ASD. As such, we believe that these definitions may provide enough information to propose quantifiable definitions that do not overly rely on human interpretation. If this is the case, it should be possible to apply computational and technological methods to their identification. Our discussion will revolve around which ASD behaviours can be considered overtly observable and can thus be identified with minimal or no reliance on human interpretation. In other words, we identify behaviours that can be tracked, measured and described by technological means.

The restrictive nature of diagnostic settings and the fact that many of the characteristics of ASD are defined by their persistence across time and different interactions (hereafter: "persistent behaviours"; e.g., "reduced sharing of interests" [14] would need to be present across multiple interactions) poses problems for temporally confined diagnostic sessions. To overcome these problems, many diagnostic tools require clinicians to observe and make judgements based on behaviours that are associated with these persistent behavioural traits (hereafter: "indicative behaviours"). For example, it was found that impairments in the perception of facial and body gestures is related to, and may be the foundation of, difficulties in social communication and intention understanding [32]. Similarly, abnormal visual processing of social information from faces [33] and impairments in visual engagement [34] have been linked with difficulties in understanding others' emotions. Evidence for such links allows diagnostic tools to use more common behaviours that do not need to be observed over time as indicators of ASD characteristics. Because persistent behaviours often require human interpretation, we argue that indicative behaviours are more appropriate as the targets for computational and technical measurement techniques. We will therefore be looking primarily at indicative behaviours, which are used by diagnostic tools and can be considered overt.

In terms of the behaviours defined by the DSM, Tables 1–3 below present an illustration of some of the considerations one must take into account when deciding the appropriateness of technologies for diagnostic purposes. In Tables 1 and 2 we identified whether each behaviour can be considered "covert" (i.e., requiring human interpretation to recognize). Those behaviours not marked as "covert" can be considered "overt". This judgement was made based on whether the behaviour can be clearly and unambiguously identified from observable behaviours alone, without having to incorporate information about the underlying intention or the appropriateness of the action. We also considered the locus of interactivity for each of the behaviours such that they are either "Interaction-Centred" (marked in Tables 1 and 2) or "Child-Centred" (not marked). Child-centred criteria are those for which only the behaviour of the child needs to be considered, for example, all the criteria under B4 (see Tables 1 and 2). Conversely, items such as all of A1 require the sensing of both interaction parties to provide an accurate assessment. These are therefore interaction-centred and impose additional challenges for automated methods; at a minimum, both the child and the therapist need to be detected and tracked by the sensory apparatus to capture the information necessary to characterise interaction-centred behaviours. It is important to note that we provide Tables 1 and 2 as a framework to illustrate the ideas presented in this review. Rather than being an authoritative classification of diagnostic criteria, we present it as a guide for future research, which should explore the viability of such applications of technology, the validity of the definitions it presents, and the development of technologies appropriate to augment the identification of each behaviour.

Table 1. Detailed breakdown of the behavioural cues for Category A that a therapist might use in ASD diagnosis based on DSM-5 criteria, and the corresponding required modalities.

	Required Modalities									**Class.**	
Behavioural cue	Gaze tracking	Speech detection	Speech analysis	Posture tracking	Gesture tracking	Facial expressions	Object tracking	Sound detection	Specific events	Covert behaviour	Interaction-centred
Category A **Persistent deficits in social communication and social interaction across contexts** **A1 Deficits in social-emotional reciprocity**											
1. One-sided conversations		✓									✓
2. Failure to offer comfort to others or to ask for it when needed			✓		✓				✓	✓	✓
3. Does not initiate conversation with peers		✓	✓	✓						✓	✓
4. Lack of showing, bringing, or pointing out objects of interest to other people				✓	✓		✓			✓	✓
5. Use of others as tools				✓	✓						✓
6. Failure to engage in simple social games				✓	✓					✓	✓
A2 Deficits in nonverbal communicative behaviours used for social interaction											
1. Impairments in social use of eye contact	✓										✓
2. Limited communication of own affect		✓	✓		✓	✓				✓	
3. Abnormalities in the use and understanding of emotion				✓	✓	✓				✓	✓
4. Impairment in the use of gestures					✓						
5. Abnormal volume, pitch, intonation, rate, rhythm, stress, prosody or volume in speech		✓									
6. Lack of coordinated verbal and nonverbal communication	✓	✓	✓		✓					✓	
A3 Deficits in nonverbal communicative behaviours used for social interaction											
1. Lacks understanding of the conventions of social interaction		✓			✓					✓	✓
2. Limited interaction with others in discussions and play	✓		✓	✓						✓	✓
3. Limited interests in talking with others			✓							✓	
4. Prefers solitary activities				✓	✓					✓	✓
5. Limited recognition of social emotions	✓	✓				✓					✓

Table 2. Detailed breakdown of the behavioural cues for Category B that a therapist might use in ASD diagnosis based on DSM-5 criteria, and the corresponding required modalities.

	Required Modalities										**Class.**
Behavioural cue	Gaze tracking	Speech detection	Speech analysis	Posture tracking	Gesture tracking	Facial expressions	Object tracking	Sound detection	Specific events	Covert behaviour	Interaction-centred
Category B **Restricted, repetitive patterns of behaviour, interests, or activities as manifested** **B1 Stereotyped or repetitive speech, motor movements, or use of objects**											
1. Repetitive hand movements					✓						
2. Stereotyped or complex whole body movements				✓							
3. Repetitive vocalisations such as repetitive guttural sounds, intonational noise making, unusual squealing repetitive humming		✓									
4. Perseverative or repetitive action / play / behaviour				✓	✓		✓				
5. Pedantic speech or unusually formal language			✓							✓	
B2 Excessive adherence to routines, ritualised patterns of verbal or nonverbal behaviour, or excessive resistance to change											
1. Overreactions to changes			✓		✓	✓				✓	✓
2. Unusual routines					✓		✓			✓	
3. Repetitive questioning about a particular topic			✓							✓	
4. Compulsions				✓	✓					✓	
B3 Highly restricted, fixated interests that are abnormal in intensity or focus											
1. Focused on the same few objects, topics or activities	✓	✓		✓	✓		✓			✓	
2. Verbal rituals		✓	✓							✓	
3. Excessive focus on irrelevant or non-functional parts of objects	✓		✓		✓		✓			✓	
B4 Hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of environment											
1. Abnormal responses to sensory input				✓		✓			✓	✓	
2. Repetitively putting hands over ears				✓				✓			
3. Extreme interest or fascination with watching movement of other things				✓	✓		✓			✓	
4. Close visual inspection of objects				✓	✓		✓				

3. Automatic Quantification of Behaviour

Some of these behaviours, as described by the DSM, are not necessarily observable; however, they are associated with indicative behaviours. For the purposes of this review, we will present the case for the observability of both indicative and DSM defined behaviours. The following discussion reviews the challenges and opportunities associated with technologies that can be used to measure behavioural modalities associated with ASD symptoms. Examples of how these technologies have or could be applied are also discussed but it is important to note that not all applications or technologies will be discussed herein; rather it is a review of the behavioural modalities which have been addressed by technologies and are relevant for ASD diagnosis. Additionally, several technologies have already been applied to therapeutic settings [35,36], or to assist individuals with ASD in their daily lives [37,38] and may be mentioned in this paper but with the view to repurposing them for diagnostic scenarios.

Similarly, since we argue that the diagnostic requirements match onto general requirements for social robotics, there is also a substantial body of literature on identifying internal states (such as emotions) from observable behaviours in more general terms. Here, we briefly discuss such relevant work where applicable before moving on to the diagnostic requirements to highlight this connection.

Finally, this is primarily an overview of the challenges and opportunities available to researchers and clinicians in this field of research, rather than a review of all research pertaining to how technologies are relevant to individuals with ASD, as such there is a substantial pool of research which is not incorporated into this discussion.

Table 3. The number of times the behaviour modalities are identified in the behavioural cues listed in Tables 1 and 2, split according to whether the behavioural cues can be considered Overt or Covert and Child-Centred or Interaction-Centred. Highlighted (in grey) cells indicate where either overt/covert or child-centred/interaction-centred are more than double its counterpart. This is on the understanding (see text) that covert cues are more difficult to automatically characterise than overt cues, and that interaction-centred cues are more (practically) difficult to assess than child-centred cues.

Modality	Total Number	Interpretability of Behaviour		Locus of Interaction		A Cues	B Cues
		Overt	Covert	Child-Centred	Interaction-Centred		
1. Gaze tracking	6	1	5	3	3	4	2
2. Speech detection	10	4	6	6	4	7	3
3. Speech Analysis	11	0	11	7	4	6	5
4. Posture tracking	15	5	10	8	7	7	8
5. Gesture tracking	19	14	5	11	8	10	9
6. Facial expressions	5	1	4	2	3	3	2
7. Object tracking	7	2	5	6	1	1	6
8. Sound detection	1	1	0	1	0	0	1
9. Specific events	2	0	2	1	1	1	1
	Total	28	48	45	31		

3.1. Gaze Behaviour

3.1.1. Intention Recognition in Social Robotics

There is already a rich pool of research applying gaze-tracking techniques to the identification of socially relevant signals. For example, Nakano and Ishii [39] used gaze information, measured using a remote eye-tracking system, to estimate how engaged a user was in a conversation with a robotic agent. Similarly, Morency and colleagues [40] trained a robotic agent to recognize whether a human interaction partner was thinking about a response or waiting for the agent to respond based on gaze behaviour. As we will see, gaze tracking with ASD populations is largely used to identify atypical gaze behaviours, rather than to interpret internal states. However, based on these findings, gaze tracking might also be useful for identifying diagnostically relevant behavioural cues such as one-sided conversations (see Table 1). That is, application of a system such as that developed by Morency and colleagues [40] could provide a quantification of how frequently a child with ASD provides a turn-taking cue, and thereby a clearer understanding of how 'one-sided' their conversation is.

3.1.2. Requirements for ASD Diagnosis

Two aspects of gaze can be tracked using technologies: head direction (which overlaps with posture detection) and eye-gaze. Head direction tracking is relatively robust, and with several readily available algorithms, (e.g., [41]). Eye-gaze tracking, however, provides a much better indication of the orientation of visual attention. The usefulness of gaze tracking in the assessment of ASD symptoms is well established. We identify gaze tracking as a potential method for assessing six of the DSM

defined behaviours (see Table 3). Additionally, studies found associations between gaze behaviours and a variety of ASD symptoms, thus demonstrating the applicability of these technologies to ASD diagnosis. For example, the absence of preferential eye-contact with approaching adults is a predictor of the level of social disability [42], and children with ASD preferentially orient visually to non-social contingencies rather than to biological motion [43]. We will focus this discussion on two types or categories of gaze tracking technology: remote systems and wearables.

The term "remote systems" here refers to any non-invasive video-based camera or system, which can be positioned in an environment to track the eye movements of participants within its field of view. These systems are perhaps most useful for measuring interaction-centred behaviours where the full social scene must be taken into account, e.g., the position of objects of interest, or of other humans. For example, joint attention tasks can only be assessed by knowing the location and direction of gaze of the interaction partners, and the position of an object to which both partners should be attending. Joint attention in particular has been noted as an area where children with ASD demonstrate atypical gaze behaviours. For instance, Swanson and Siller [44] examined whether there were differences in the gaze behaviours of typically developing (TD) and ASD children during a joint attention task. They used a single remote system attached to a computer screen that displayed videos of an actor. Children's gaze behaviours were measured while they watched the video to see if they attended to the same areas of the screen as the actor. While Swanson and Siller did not find any differences between groups in global measures of gaze (e.g., overall looking time), they did detect differences in the microstructure of gaze behaviour (e.g., duration of first fixation). This not only demonstrates that gaze tracking is useful in the assessment of ASD behaviours, but also that using such technologies can allow us to identify behaviours which may not be identified by human observers.

Wearable gaze tracking systems range from head-mounted cameras to eye-tracking glasses and can be worn either by the child undergoing assessment or by a clinician or parent who is interacting with the child. Wearables allow the wearer more freedom of movement than remote systems and can be implemented outside of the diagnostic setting, allowing clinicians to gather diagnostic information about the child's daily life and at-home behaviours. Wearables are more appropriate for examining precisely what a child is looking at, i.e., investigations of attention orienting, in more naturalistic or dynamic settings. For example, Magrelli et al. [45] investigated how TD children and children with ASD orient their attention to social stimuli using a head-mounted eye-tracking device. This study specifically examined child behaviour during dyadic play interactions with an adult in environments that were familiar to the children. Magrelli et al. found that children with ASD looked at the adult's face less than TD children. This study demonstrates how wearable eye-tracking technologies could allow ASD diagnosis to include empirical, quantitative data about the child's behaviour during their every-day lives.

However, each of these techniques is associated with several challenges when applied to diagnostic settings and, therefore, opportunities for future development. For instance, the use of remote cameras requires some amount of restriction to the child's movements. To provide a full-frontal view of the face, single-camera techniques require the child to be relatively stationary and are ideally implemented to assess a child's behaviour during a task tailored to elicit differential eye-movements in ASD and TD children (as in [46]). Diagnostic settings however, often involve engaging children in several different tasks to assess a range of behaviours. Techniques such as switching between multiple cameras to find the optimal view seem, therefore, more appropriate to this setting. Wearables also offer a solution to this problem; however, the need for compact and comfortable technologies often results in some loss to the technology's accuracy [47].

3.2. Speech Behaviour

3.2.1. Intention Recognition in Social Robotics

It is well established that internal states and social signals can be recognized from features of speech. In particular, emotional states such as happiness, sadness, anger and fear were classified based on prosodic features of speech [48–50]. Similarly, prosodic features have been used to train classifiers to distinguish between positive, negative and neutral emotional states [51]. In terms of social signals, Hsiao et al. [52], for instance, demonstrated that turn-taking patterns and prosody features in speech could be used to classify high and low social engagement. This evidence clearly demonstrates that internal state information and social signals can be identified by classification systems based on speech and verbal behaviours.

3.2.2. Requirements for ASD Diagnosis

Speech processing has received increasing attention in recent years as commercial applications have come to the public. Solutions therefore exist that could be applied to automated analysis of speech during general, as well as diagnostic, interactions [53], although variability between speakers poses problems [54] that are particularly acute with child voices [55,56]. There are two broad types of speech properties that may be distinguished in the context of the diagnostic criteria: (1) detection of the presence/absence of speech (10 criteria; Table 3); and (2) the processing of the content of speech (comprised of detection of reportative speech, keyword recognition and understanding – 11 criteria; Table 3). The first of these can be addressed through the application of statistically-based signal processing techniques, for which there are a range of established solutions (e.g., [57,58]). Keyword recognition (which could also be used for repetition detection) lies in the area of speech recognition that is similarly well supported by a range of methods [57], including deep learning systems [59], although the complexity and noisiness of real-world contexts present further limitations. Speech understanding poses the most challenging level of analysis, with current technologies being limited to constrained settings until a greater level of context information can be incorporated [60]. In all of these cases, maximising the quality of the sound recordings using microphones (while minimising background noise, interference, etc) is clearly beneficial for maximising the performance of automated methods. In application to the diagnosis of ASD this may necessitate the deployment of multiple microphones, which introduces further issues of signal integration and sound source localisation, particularly with multiple speakers (e.g., the child and the clinician) present [61].

Children with ASD have difficulties both in generating and recognising vocal prosody and intonation [62], display a deficit in syllable production [63], and have substantially higher proportions of atypical vocalizations than TD children [64]. Differences in communication tend to be persistent, show little change over time, and may include monotonic intonation, deficits in the use of pitch and control of volume, in vocal quality, and use of aberrant stress patterns [16,65]. All these patterns can be observed around the age of 2, which has been proposed as the age at which a reliable diagnosis can be provided [66]. We identified a total of seventeen diagnostic behaviours as observable via speech behaviours (Tables 1 and 2). One of the main benefits of automated speech analysis for ASD diagnosis is that its use could speed up the assessment process in that clinicians would not be required to listen to and hand-code recordings of child speech. The second advantage we consider is that the use of technology allows for the assessment of child speech in their everyday lives and naturalistic interactions. For example, Warren et al. [67] used a digital language processor and language analysis software to record and analyse the conversational environments of children with ASD and TD children. The children wore the recording equipment in a pocket of their own clothing. They found that children with ASD engaged in fewer conversations and produced fewer vocalisations than TD children. Additionally, Warren et al. were able to examine what effect the language use and skills of the adults in the children's environments had on child speech. Their analysis of this data showed that the different language environments provided by adults (e.g., number of different words produced by adults,

frequency of responses to child utterances) may influence a child's linguistic development and thereby impose confounds into assessments of speech in children with ASD. While this technology can also be implemented within a classical diagnostic setting, this study demonstrates some of the benefits of technologies for gathering naturalistic data for assessment, which includes obtaining data that might otherwise be unavailable to clinicians (i.e., the child's language environment).

3.3. Posture and Gesture Behaviour

3.3.1. Intention Recognition in Social Robotics

Vision-based methods (using standard cameras/2D images) for human motion capture are well established [68], with face tracking being particularly developed. The recent advent of depth-based tracking and processing of detected skeletons in the scene (primarily using RGB-D data) resulted in additional well-established tools to facilitate various types of pose and behaviour analysis [69]. Depth-based methods can also be applied to hand-gesture characterisation [70], although sensory resolution constraints (e.g., hands and fingers being more difficult to detect) mean that image-based methods may currently remain more appropriate [71].

There is evidence demonstrating that emotional states (e.g., happiness, sadness, anger) [72–74] and internal states such as engagement [75] can be recognised from gesture and posture information collected through standard digital video devices. Similarly, body postures captured using the Microsoft Xbox Kinect device were successfully used to classify emotional states [76]. Outside of emotion recognition, other research showed that internal states and socially relevant dispositions or states can be recognised through pose and gestures. Okada et al. [77] were able to classify dominance and leadership based on gesture information. The main concern for using gesture and posture information during human-robot interactions "in the wild" is that fitting a robotic agent with a camera suitable for this purpose is not always straightforward. Current research generally relies on being able to use a camera system separate from any robotic agent, thus restricting the interaction environment. This is not to say that it is not achievable. Ramey and colleagues [78] for example, integrated the Kinect device into a social robot for tracking and recognising hand gestures. Similarly, Elfaramawy and colleagues [74] mounted a depth sensor onto a Nao robot to record movement data during an interaction with human users. This data was then used to classify whether the interaction partner was expressing the emotions anger, fear, happiness, sadness or surprise. These results demonstrate that internal state information and socially relevant information can be interpreted from gesture and posture behaviours.

3.3.2. Requirements for ASD Diagnosis

In terms of information directly relevant to the diagnostic criteria, methods of tracking and recognizing posture and gesture behaviours are typically targeted at the characterisation of individuals rather than groups of people, and so would be most appropriate for overt and child-centred behaviours, followed by overt and interaction-centred behaviours, provided both parties in the interaction are tracked. Twenty-four of the behaviours in Tables 1 and 2 are observable via posture and/or gesture behaviours.

Many of these behaviours are captured by research exploring deficits in motor-skills. The developmental trajectory of motor skills has been demonstrated to be predictive of the rate of language development [79,80], deficits in adaptive behaviour skills [81] and social communication skills [82]. Some studies conclude that between 80–90% of children with ASD show some degree of impairment in motor skills [83,84], and a recent meta-analysis concluded that motor deficits should be included in the core symptoms of ASD [85]. Furthermore, deficits in motor skills may affect fine and gross motor coordination, stereotyped movements and awkward patterns of object manipulation, lack of purposeful exploratory movements, and alterations of movement planning and execution [86–88]. Cook and colleagues [89] used a motion tracking system to explore whether individuals with ASD

demonstrated atypical kinematic profiles in arm movements compared to TD individuals. They found that individuals with ASD produced arm movements that were jerkier and proceeded with greater acceleration and velocity. Similarly, Anzulewicz et al. [90] used the sensors available in an iPad mini to measure the motor activity displayed by children with ASD as they played games on the device. Machine learning analysis of this data was used to identify whether there were differences between children with ASD and TD children, and found that children with ASD exhibited greater force of contact, different distributions of forces within gestures, and differences in gesture kinematics. Together these studies demonstrate not only that diagnostic information is available in behaviours which can be measured via motion sensing technologies, but also that these technologies are readily available in smart devices such as tablets and other touch screens.

Most demonstrations of technologies measuring atypical postures and gestures produced by individuals with ASD involve choreographed or specific motions and tasks (e.g., [89]). As such, more data of naturalistic gestures may be required before this technology can be fully implemented in diagnostic settings. The goal would be to provide data describing the differences between children with ASD and TD children in the kinds of gestures that are produced in social interactions and within the tasks involved in diagnostic assessments. However, with such a dataset, motion tracking technologies have a great potential for augmenting the diagnostic process by providing clinicians with information which is difficult to assess by human observers but which contains diagnostic identifiers.

3.4. Object and Sound Detection

Seven of the behaviours in Tables 1 and 2 also require object tracking and one requires sound detection. These modalities are considered separately from those in the paragraphs above since they are not directed specifically at a human agent. However, the same set of sensors may be deployed as for the other behavioural modalities, namely cameras (using 2D and depth images) and microphones.

Object tracking is particularly useful for assessments of joint attention, and in the ways children with ASD attend to and express their interest in objects. For example, Elison et al. [91] were able to categorise the behaviours of 12-month old children into distinct groups based on observed repetitive object manipulation behaviours. Furthermore, those children who demonstrate more repetitive object manipulation behaviours were more likely to be diagnosed with ASD at 24 months. Automating the measurement of these behaviours would require both gesture and object tracking but could reveal further identifiers for ASD or allow us to more precisely quantify the differences between groups on this type of task. Most demonstrations of automated object tracking in ASD contexts come in the form of robot-assisted therapies or diagnostic protocols. Petric et al. [92], for example, tested the efficacy of their autonomous robot protocol in carrying out four diagnostic tasks with children. In relation to object-tracking, these tasks involved the robot detecting whether the child was playing with a toy before attracting the child's attention (response to name), directing a child's attention to an object (joint attention), and to test whether a child would imitate actions using functional objects (functional imitation). The systems implemented in this study involved both the tracking of objects and the assessment of the child's behaviour with or towards that object in real time. While this application of object-tracking technologies is different to the application we propose in this review (i.e., we are not necessarily proposing the use of robots), this study does demonstrate how object tracking, alongside other methods like gesture tracking, can be used to assess child behaviour in real time during a clinical assessment to provide useful feedback.

There are a range of well-established methods and algorithms in the literature that are effective for object tracking based on visual data, with recent advances using deep learning methods (e.g., [93]). However, if manipulation is involved (as in items B1.4 and B3.1), then object occlusions may be problematic and so should be a focus of future developments. An additional challenge to this technique is that there is little empirical work quantifying differences between how children with ASD and TD children manipulate objects. Such work is essential before these techniques can be implemented in a

diagnostic setting because it would provide us with the identifiers, if there are any, which can be used to distinguish between children with and without ASD.

3.5. Facial Expressions

3.5.1. Intention Recognition in Social Robotics

Numerous technologies and approaches were developed to recognise and classify emotional facial expressions (EFEs). It has been demonstrated that emotional states can be recognised from facial expressions extracted from video data [94–98], (see also [99] for a survey of methods). Facial expressions have also proved useful for classifying engagement [100,101] showing that facial expressions are useful for identifying social signals beyond emotions.

3.5.2. Requirements for ASD Diagnosis

While the symptoms involving emotion expression have all been categorised as covert or "requiring human interpretation", technologies and techniques for identifying facial expressions, such as those described above, would be helpful in the assessment of how children communicate their own emotional states. However, this would be limited to examining the "strength" or frequency of emotional facial expressions rather than their appropriateness as this element requires human interpretation. Additionally, emotional expression analysis could aid in assessing how children detect and respond to the emotional expressions of others by combining such methods with gesture or eye tracking, or speech analysis. One study found that typically developed participants demonstrate different fixation and scanning patterns when observing faces expressing different emotions (e.g., more gazing at the mouth for happy and angry faces, and the eyes for sad faces) [102]. Additionally, another study found that children with ASD fixated on the mouth of happy and angry faces less than their TD peers [103]. If we take these findings together, they demonstrate a use-case for technologies which can be applied in naturalistic settings and are capable of simultaneously tracking the emotional expressions being communicated towards a child, and the child's gaze behaviours in viewing those expressions. This application would allow clinicians to include naturalistic data on emotion recognition capabilities in their diagnostic analysis. Alternatively, if this same method were applied in a controlled clinical setting, the use of automated emotion recognition would firstly help in validating whether an emotional expression was sufficient to communicate one emotion over another. Additionally, it would reduce the time needed to assess a child's gaze behaviours by automating the mapping between the occurrence of an emotional expression and the child's gaze behaviours in processing this expression, thus eliminating the need to manually code and map these events together.

Automated emotion classification from faces is typically based on the six basic emotions [104], and are associated with numerous limitations when applied to real-world situations (see [99,105] for reviews). However, given that during a diagnostic assessment, the clinician would act out the emotional expression (thus exaggerating the features), such methods may nevertheless be appropriate. Classification methods typically use Action Unit coding of facial expression features, with more recent attempts to incorporate other visual information, such as head behaviour [106]. Being a camera-based method, this characterisation of facial expression is subject to similar constraints as posture and gaze analysis.

4. Discussion and Conclusions

4.1. Limitations of Current Technology

In this paper, we discussed the state of the art of technological means to measure behavioural cues relevant to the diagnostic criteria for ASD. A consistent and reliable quantification of behaviour in the modalities identified that would go beyond the observational techniques currently employed has the potential to present clear advantages to clinicians in their evaluation of ASD symptoms.

It is apparent from our review that while there is definite scope for such automated quantification, there remain several limitations with current sensory technologies and their associated methods in this context. Some are due to practical constraints (e.g., the positioning and coverage of individual sensors), but the more problematic issues are typically related to diagnostic criteria involving a covert behavioural component, i.e., those behaviours that require some degree of interpretation in addition to the observation of the overt phenomena. Human assessors naturally bring their prior experience and extensive training into the diagnostic assessment process; for automated methods, this prior knowledge and experience must be codified for it to be applied. The problematic qualitative nature of such developed experience is an area in which the sensory interpretation methods discussed are currently lacking, for which deeper, more complex (perhaps even cognitive) models are required if they are to be sufficient to adequately augment human characterisation efforts.

Work in this direction must start on the more general level, outside of the confines of therapeutic settings. We have highlighted several existing works demonstrating how covert states/behaviours may be identified from overt behavioural cues at this level. A large body of work, for example, is devoted to the recognition of emotional states in a range of contexts. However, this is usually limited to the six 'basic' emotions [104] or to identifying the valence of emotion (positive, negative or neutral). As such, more work in this area is needed. In particular, further explorations of whether different, more complex covert states (e.g., frustration, distress, confusion) are shown in overt behaviours.

4.2. Classes of Behavioural Modalities in ASD Diagnosis

Seven behavioural modalities were described, which can be considered overt and therefore identifiable via technological means. Additionally, Tables 1 and 2 provide an initial framework for deciding which modalities are most appropriate for identifying and tracking these diagnostic behaviours. We propose this framework as a guideline for clinicians wishing to incorporate technological means of behaviour measurement into the diagnosis of ASD, as well as for researchers looking to develop and improve such technologies. In addressing the former goal, we have also identified behaviours we believe to be mostly, if not entirely, overtly observable. While covert behaviours do pose a challenge to technological measurement techniques, due to the requirement for human interpretation, our review identified some overt behaviours that were shown to be associated with, or indicative of, some of these behaviours. As such, the technologies and approaches we have discussed present an opportunity for clinicians to demonstrate support for their observations using quantifiable behaviours. For example, in assessing a child's ability to recognise emotional facial expressions, clinicians could both observe children's reactions to such expressions and measure the child's gaze patterns. This would not only provide empirical support for the clinician's conclusion, but may also assist in disambiguating a child's behaviour where there is uncertainty.

Alongside the overtness of each behaviour, we have also distinguished between behaviours that are expressed solely by the child being assessed (Child-Centred) and which are uniquely expressed within an interaction (Interaction-Centred). This distinction provides a framework for deciding which technologies or set-ups are most appropriate for measuring each behaviour, e.g., is a single camera more appropriate than multiple cameras (capturing the behaviour of all members of the interaction) for collecting visual data about a joint-attention assessment? Interaction-Centred behavioural cues do present complications in that they entail the tracking and characterisation of multiple individuals (minimally the child and the clinician) and their coordination, which is feasible, though posing additional challenges. Accounting for these considerations, it is noticeable that some of the modalities lend themselves more readily to immediate application than others, gesture tracking being the clearest example of this. Conversely, speech analysis remains a challenge, even assuming high performing speech recognition. Furthermore, we observe that 63% of behavioural cues across modalities require some degree of interpretation, and which would thus be currently difficult to automate.

4.3. Future Work

4.3.1. Diagnosis of ASD

Existing studies that deal with the use of technology in the diagnosis or treatment of ASD emphasise methodological differences in this broad field [107]. Our review suggests that more effort should be invested in developing technology-based applications that aim to benefit the diagnostic process for children with developmental disabilities, such as ASD or ADHD [108]. An additional, perhaps even greater, challenge in this field is not just to create effective technologies, but also to make them accessible for practitioners in terms of availability, ease of operation and cost. Technology-based tools have the potential to be an important resource in both assessment and treatment for individuals with ASD as they may be able to reduce the time and effort required by expert clinicians. As a result, diagnoses would become more accessible, consistent (through the application of standard recognition technologies for those overt aspects), and, potentially, more understandable. For instance, if a caregiver understands that a child's difficulty with recognising emotional facial expressions is related to the way the child attends to different facial features, the caregiver is able to apply this knowledge when providing the child with support during their daily lives, e.g., overtly directing the child's attention to relevant features during emotion-recognition games/exercises.

4.3.2. Social Robotics

As far as the field of social robotics is concerned, we have highlighted the need for algorithms that can infer covert, or internal states from observable kinematics. We have shown, in particular, that the main limitation is primarily on the algorithmic side and we recommend that more effort is put on addressing this directly. Indeed, we suggest (Section 4.1) that it may be necessary to integrate a more general cognitive aspect to this algorithmic processing. This provides a motivation for consideration of cognitive architectures in social robotics [109]: as we have highlighted in this paper, a robot controller that is merely responsive to observable behaviour is very unlikely to be sufficient for autonomous social interaction. As a means to further research in this direction, we have highlighted the overlap between the requirements of social robotics in general and ASD diagnosis in particular: as such, we argue that a system which can satisfactorily address the latter will also contain the technological developments required to advance the former.

4.4. Conclusion

Overall, this contribution highlighted that we are now at a point where it is feasible to incorporate novel, technology-based means into the diagnostic process for ASD. This opens up a new avenue of research, now ripe for exploring, focused on thorough evaluations of the benefits of, and further challenges in, technology-augmented diagnosis. With this paper, we hope to have provided the necessary starting points, highlighting for clinicians what is already possible, and for the developers of technology and psychology researchers, what the immediate obstacles are from a diagnostic point of view. The intent is to provide reliable and consistent quantitative data with which the diagnostic process can be improved, resulting in positive impacts for those children concerned. At the same time, it also highlights that further development of algorithms that can suitably assess covert states is a research avenue ready to be explored further in social robotics in general: with technological issues mostly solved and a good understanding of human-robot interactions from Wizard-of-Oz studies, this is the missing piece of the puzzle.

Author Contributions: Conceptualization, C.C., P.B. and S.T.; Formal analysis, M.E.B., C.C., P.B. and S.T.; Funding acquisition, C.C. and S.T.; Investigation, M.E.B., C.C., P.B. and S.T.; Methodology, M.E.B., C.C., P.B. and S.T.; Supervision, S.T.; Visualization, P.B. and S.T.; Writing—original draft, M.E.B., C.C. and P.B.; Writing— review & editing, P.B. and S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Commission FP7 grant number 611391.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Carruthers, P.; Smith, P.K. *Theories of Theories of Mind*; Cambridge University Press: Cambridge, UK, 1996.
2. Svensson, H.; Thill, S. Beyond bodily anticipation: internal simulations in social interaction. *Cognit. Syst. Res.* **2016**, *40*, 161–171. [CrossRef]
3. Demiris, Y.; Khadhouri, B. Hierarchical attentive multiple models for execution and recognition of actions. *Robot. Auton. Syst.* **2006**, *54*, 361–369. [CrossRef]
4. Demiris, Y. Prediction of intent in robotics and multi-agent systems. *Cognit. Process.* **2007**, *8*, 151–158. [CrossRef] [PubMed]
5. Haruno, M.; Wolpert, D.M.; Kawato, M. MOSAIC Model for Sensorimotor Learning and Control. *Neural Comput.* **2001**, *13*, 2201–2220. [CrossRef]
6. Metta, G.; Sandini, G.; Natale, L.; Craighero, L.; Fadiga, L. Understanding mirror neurons: A bio-robotic approach. *Interact. Stud.* **2006**, *7*, 197–232. [CrossRef]
7. Bartlett, M.E.; Edmunds, C.E.R.; Belpaeme, T.; Thill, S.; Lemaignan, S. What Can You See? Identifying Cues on Internal States From the Movements of Natural Social Interactions. *Front. Robot. AI* **2019**, *6*, 49. [CrossRef]
8. Bradwell, H.L.; Edwards, K.J.; Winnington, R.; Thill, S.; Jones, R.B. Companion robots for older people: importance of user-centred design demonstrated through observations and focus groups comparing preferences of older people and roboticists in South West England. *BMJ Open* **2019**. [CrossRef]
9. Vollmer, A.L.; Read, R.; Trippas, D.; Belpaeme, T. Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Sci. Robot.* **2018**. [CrossRef]
10. Belpaeme, T.; Kennedy, J.; Ramachandran, A.; Scassellati, B.; Tanaka, F. Social robots for education: A review. *Sci. Robot.* **2018**. [CrossRef]
11. Cao, H.; Esteban, P.G.; Bartlett, M.; Baxter, P.; Belpaeme, T.; Billing, E.; Cai, H.; Coeckelbergh, M.; Costescu, C.; David, D.; et al. Robot-Enhanced Therapy: Development and Validation of Supervised Autonomous Robotic System for Autism Spectrum Disorders Therapy. *IEEE Robot. Autom. Mag.* **2019**, *26*, 49–58. [CrossRef]
12. Dautenhahn, K.; Werry, I. Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmat. Cognit.* **2004**, *12*, 1–35. [CrossRef]
13. Scassellati, B.; Admoni, H.; Matarić, M. Robots for Use in Autism Research. *Annu. Rev. Biomed. Eng.* **2012**, *14*, 275–294. [CrossRef] [PubMed]
14. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*; American Psychiatric Publishing: Washington, DC, USA, 2013.
15. Kanner, L. Autistic disturbances of affective contact. *Nerv. Child* **1943**, *2*, 217–250.
16. Kanner, L. Follow-up study of eleven autistic children originally reported in 1943. *J. Autism Child. Schizophr.* **1971**, *1*, 119–145. [CrossRef]
17. Grzadzinski, R.; Huerta, M.; Lord, C. DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. *Mol. Autism* **2013**, *4*, 12. [CrossRef]
18. Huerta, M.; Lord, C. Diagnostic Evaluation of Autism Spectrum Disorders. *Pediatr. Clin. N. Am.* **2012**, *59*, 103–111. [CrossRef]
19. Yates, K.; Le Couteur, A. Diagnosing autism. *Paediatr. Child Health* **2013**, *23*, 5–10. [CrossRef]
20. Rogers, C.L.; Goddard, L.; Hill, E.L.; Henry, L.A.; Crane, L. Experiences of diagnosing autism spectrum disorder: A survey of professionals in the United Kingdom. *Autism* **2016**, *20*, 820–831. [CrossRef]
21. Le Couteur, A.; Lord, C.; Rutter, M. *The Autism Diagnostic Interview-Revised (ADI-R)*; Western Psychological Services: Los Angeles, CA, USA, 2003.
22. Schopler, E.; Reichler, R.J.; DeVellis, R.F.; Daly, K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *J. Autism Dev. Disord.* **1980**, *10*, 91–103. [CrossRef]
23. Lord, C.; Risi, S.; Lambrecht, L.; Edwin H. Cook, J.; Leventhal, B.L.; DiLavore, P.C.; Pickles, A.; Rutter, M. The Autism Diagnostic Observation Schedule–Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *J. Autism Dev. Disord.* **2000**, *30*, 205–223. [CrossRef]

24. Zwaigenbaum, L.; Bryson, S.; Garon, N. Early identification of autism spectrum disorders. *Behav. Brain Res.* **2013**, *251*, 133–146. [CrossRef] [PubMed]
25. Falkmer, T.; Anderson, K.; Falkmer, M.; Horlin, C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. *Eur. Child Adolesc. Psychiatry* **2013**, *22*, 329–340. [CrossRef] [PubMed]
26. Crane, L.; Chester, J.W.; Goddard, L.; Henry, L.A.; Hill, E. Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism* **2015**, *20*, 153–162. [CrossRef]
27. Doi, K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput. Med Imaging Graph.* **2007**, *31*, 198–211. [CrossRef] [PubMed]
28. Xiao, Y.; Zeng, J.; Niu, L.; Zeng, Q.; Wu, T.; Wang, C.; Zheng, R.; Zheng, H. Computer-Aided Diagnosis Based on Quantitative Elastographic Features with Supersonic Shear Wave Imaging. *Ultrasound Med. Biol.* **2014**, *40*, 275–286. [CrossRef]
29. Hallgren, K.A. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor. Quant. Methods Psychol.* **2012**, *8*, 23–34. [CrossRef]
30. Zander, E.; Willfors, C.; Berggren, S.; Coco, C.; Holm, A.; Jifält, I.; Kosieradzki, R.; Linder, J.; Nordin, V.; Olafsdottir, K.; Bölte, S. The Interrater Reliability of the Autism Diagnostic Interview-Revised (ADI-R) in Clinical Settings. *Psychopathol.* **2017**, *50*, 219–227. [CrossRef]
31. Zander, E.; Willfors, C.; Berggren, S.; Choque-Olsson, N.; Coco, C.; Elmund, A.; Moretti, Å.H.; Holm, A.; Jifält, I.; Kosieradzki, R.; Linder, J.; Nordin, V.; Olafsdottir, K.; Poltrago, L.; Bölte, S. The objectivity of the Autism Diagnostic Observation Schedule (ADOS) in naturalistic clinical settings. *Eur. Child Adolesc. Psychiatry* **2015**, *25*, 769–780. [CrossRef]
32. O'Brien, J.; Spencer, J.; Girges, C.; Johnston, A.; Hill, H. Impaired Perception of Facial Motion in Autism Spectrum Disorder. *PLoS ONE* **2014**, *9*, e102173. [CrossRef]
33. Adolphs, R.; Sears, L.; Piven, J. Abnormal Processing of Social Information from Faces in Autism. *J. Cognit. Neurosci.* **2001**, *13*, 232–240. [CrossRef]
34. Sacrey, L.A.R.; Armstrong, V.L.; Bryson, S.E.; Zwaigenbaum, L. Impairments to visual disengagement in autism spectrum disorder: A review of experimental studies from infancy to adulthood. *Neurosci. Biobehav. Rev.* **2014**, *47*, 559–577. [CrossRef] [PubMed]
35. Scassellati, B.; Boccanfuso, L.; Huang, C.M.; Mademtzi, M.; Qin, M.; Salomons, N.; Ventola, P.; Shic, F. Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* **2018**, *3*, eaat7544. [CrossRef]
36. Washington, P.; Wall, D.; Voss, C.; Kline, A.; Haber, N.; Daniels, J.; Fazel, A.; De, T.; Feinstein, C.; Winograd, T. SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism. Available online: https://dl.acm.org/doi/pdf/10.1145/3130977 (accessed on 31 January 2020).
37. Gentry, T.; Kriner, R.; Sima, A.; McDonough, J.; Wehman, P. Reducing the Need for Personal Supports Among Workers with Autism Using an iPod Touch as an Assistive Technology: Delayed Randomized Control Trial. *J. Autism Dev. Disord.* **2014**, *45*, 669–684. [CrossRef] [PubMed]
38. Tang, Z.; Guo, J.; Miao, S.; Acharya, S.; Feng, J.H. Ambient Intelligence Based Context-Aware Assistive System to Improve Independence for People with Autism Spectrum Disorder. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016. [CrossRef]
39. Nakano, Y.; Ishii, R. Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In Proceedings of the 15th international conference on Intelligent user interfaces, Hong Kong, China, 7–10 February 2010; pp. 139–148.
40. Morency, L.P.; Christoudias, C.M.; Darrell, T. Recognizing gaze aversion gestures in embodied conversational discourse. In Proceedings of the 8th international conference on Multimodal interfaces, Banff, AB, Canada, 2–4 November 2006; pp. 287–294.
41. Zhu, X.; Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012. [CrossRef]
42. Jones, W.; Carr, K.; Klin, A. Absence of Preferential Looking to the Eyes of Approaching Adults Predicts Level of Social Disability in 2-Year-Old Toddlers With Autism Spectrum Disorder. *Arch. Gen. Psychiatry* **2008**, *65*, 946. [CrossRef] [PubMed]
43. Klin, A.; Lin, D.J.; Gorrindo, P.; Ramsay, G.; Jones, W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* **2009**, *459*, 257–261. [CrossRef]

44. Swanson, M.R.; Siller, M. Patterns of gaze behavior during an eye-tracking measure of joint attention in typically developing children and children with autism spectrum disorder. *Res. Autism Spectr. Disord.* **2013**, *7*, 1087–1096. [CrossRef]
45. Magrelli, S.; Jermann, P.; Noris, B.; Ansermet, F.; Hentsch, F.; Nadel, J.; Billard, A. Social orienting of children with autism to facial expressions and speech: a study with a wearable eye-tracker in naturalistic settings. *Front. Psychol.* **2013**, *4*. [CrossRef]
46. Frazier, T.W.; Klingemier, E.W.; Beukemann, M.; Speer, L.; Markowitz, L.; Parikh, S.; Wexberg, S.; Giuliano, K.; Schulte, E.; Delahunty, C.; Ahuja, V.; Eng, C.; Manos, M.J.; Hardan, A.Y.; Youngstrom, E.A.; Strauss, M.S. Development of an Objective Autism Risk Index Using Remote Eye Tracking. *J. Am. Acad. Child Adolesc. Psychiatry* **2016**, *55*, 301–309. [CrossRef]
47. Noris, B.; Nadel, J.; Barker, M.; Hadjikhani, N.; Billard, A. Investigating Gaze of Children with ASD in Naturalistic Settings. *PLoS ONE* **2012**, *7*, e44144. [CrossRef]
48. Petrushin, V.A. Emotion recognition in speech signal: experimental study, development, and application. In Proceedings of the Sixth International Conference on Spoken Language Processing, Beijing, China, 16–20 October 2000.
49. Dai, K.; Fell, H.J.; MacAuslan, J. Recognizing emotion in speech using neural networks. *Telehealth Assist. Technol.* **2008**, *31*, 38.
50. Li, Y.; Zhao, Y. Recognizing emotions in speech using short-term and long-term features. In Proceedings of the Fifth International Conference on Spoken Language Processing, Sydney, Australia, 30 November–4 December 1998.
51. Litman, D.; Forbes, K. Recognizing emotions from student speech in tutoring dialogues. In Proceedings of the 2003 IEEE Workshop on Automatic Speech Recognition and Understanding, St Thomas, VI, USA, 30 November–4 December 2003; pp. 25–30.
52. Hsiao, J.C.y.; Jih, W.r.; Hsu, J.Y.j. Recognizing continuous social engagement level in dyadic conversation by using turn-taking and speech emotion patterns. In Proceedings of the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.
53. Oller, D.K.; Niyogi, P.; Gray, S.; Richards, J.A.; Gilkerson, J.; Xu, D.; Yapanel, U.; Warren, S.F. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13354–13359. [CrossRef] [PubMed]
54. Benzeghiba, M.; Mori, R.D.; Deroo, O.; Dupont, S.; Erbes, T.; Jouvet, D.; Fissore, L.; Laface, P.; Mertins, A.; Ris, C.; Rose, R.; Tyagi, V.; Wellekens, C. Automatic speech recognition and speech variability: A review. *Speech Commun.* **2007**, *49*, 763–786. [CrossRef]
55. Gerosa, M.; Giuliani, D.; Brugnara, F. Acoustic Variability and Automatic Recognition of Children's Speech. *Speech Commun.* **2007**, *49*, 847–860. [CrossRef]
56. Kennedy, J.; Lemaignan, S.; Montassier, C.; Lavalade, P.; Irfan, B.; Papadopoulos, F.; Senft, E.; Belpaeme, T. Child Speech Recognition in Human-Robot Interaction. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017. [CrossRef]
57. Rabiner, L.R.; Schafer, R.W. Introduction to Digital Speech Processing. *Found. Trends Signal Process.* **2007**, *1*, 1–194. [CrossRef]
58. Ramirez, J.; M., J.; C., J. Voice Activity Detection. Fundamentals and Speech Recognition System Robustness. In *Robust Speech Recognition and Understanding*; I-Tech: Vienna, Austria, 2007; Volume 6, pp. 1–22. [CrossRef]
59. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Rahman Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [CrossRef]
60. Moore, R.K. Spoken language processing: Piecing together the puzzle. *Speech Commun.* **2007**, *49*, 418–435. [CrossRef]
61. Athanasopoulos, G.; Verhelst, W.; Sahli, H. Robust speaker localization for real-world robots. *Comput. Speech Lang.* **2015**, *34*, 129–153. [CrossRef]
62. Shriberg, L.D.; Paul, R.; McSweeny, J.L.; Klin, A.; Cohen, D.J.; Volkmar, F.R. Speech and Prosody Characteristics of Adolescents and Adults With High-Functioning Autism and Asperger Syndrome. *J. Speech Lang. Hear. Res.* **2001**, *44*, 1097–1115. [CrossRef]
63. Rapin, I.; Dunn, M. Language disorders in children with autism. *Semin. Pediatr. Neurol.* **1997**, *4*, 86–92. [CrossRef]

64. Sheinkopf, S.J.; Mundy, P.; Oller, D.K.; Steffens, M. Vocal Atypicalities of Preverbal Autistic Children. *J. Autism Dev. Disord.* **2000**, *30*, 345–354. [CrossRef]
65. Paul, R.; Augustyn, A.; Klin, A.; Volkmar, F.R. Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *J. Autism Dev. Disord.* **2005**, *35*, 205–220. [CrossRef] [PubMed]
66. Centres for Disease Control and Prevention. Autism Spectrum Disorder (ASD): Data and Statistics. Available online: https://www.cdc.gov/ncbddd/autism/data.html (accessed on 31 January 2020).
67. Warren, S.F.; Gilkerson, J.; Richards, J.A.; Oller, D.K.; Xu, D.; Yapanel, U.; Gray, S. What Automated Vocal Analysis Reveals About the Vocal Production and Language Learning Environment of Young Children with Autism. *J. Autism Dev. Disord.* **2009**, *40*, 555–569. [CrossRef] [PubMed]
68. Moeslund, T.B.; Hilton, A.; Krüger, V. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **2006**, *104*, 90–126. [CrossRef]
69. Han, J.; Shao, L.; Xu, D.; Shotton, J. Enhanced Computer Vision With Microsoft Kinect Sensor: A Review. *IEEE Trans. Cybern.* **2013**, *43*, 1318–1334. [CrossRef] [PubMed]
70. Suarez, J.; Murphy, R.R. Hand gesture recognition with depth images: A review. In Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, 9–13 September 2012. [CrossRef]
71. Mitra, S.; Acharya, T. Gesture Recognition: A Survey. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2007**, *37*, 311–324. [CrossRef]
72. Castellano, G.; Villalba, S.D.; Camurri, A. Recognising human emotions from body movement and gesture dynamics. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction, Lisbon, Portugal, 12–14 September 2007.
73. Saha, S.; Datta, S.; Konar, A.; Janarthanan, R. A study on emotion recognition from body gestures using Kinect sensor. In Proceedings of the 2014 International Conference on Communication and Signal Processing, Bangkok, Thailan, 10–12 October 2014; pp. 56–60.
74. Elfaramawy, N.; Barros, P.; Parisi, G.I.; Wermter, S. Emotion recognition from body expressions with a neural network architecture. In Proceedings of the 5th International Conference on Human Agent Interaction, Bielefeld, Germany, 17–20 October; pp. 143–149.
75. Sanghvi, J.; Castellano, G.; Leite, I.; Pereira, A.; McOwan, P.W.; Paiva, A. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne Switzerland, 6–9 March 2011; pp. 305–312.
76. Guerrero Rázuri, J.F.; Larsson, A.; Sundgren, D.; Bonet, I.; Moran, A. Recognition of emotions by the emotional feedback through behavioral human poses. *Int. J. Comput. Sci. Issues* **2015**, *12*, 7–17.
77. Okada, S.; Aran, O.; Gatica-Perez, D. Personality trait classification via co-occurrent multiparty multimodal event discovery. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November; pp. 15–22.
78. Ramey, A.; González-Pacheco, V.; Salichs, M.A. Integration of a low-cost RGB-D sensor in a social robot for gesture recognition. In Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Lausanne, Switzerland, 6–9 March 2011; pp. 229–230.
79. Bedford, R.; Pickles, A.; Lord, C. Early gross motor skills predict the subsequent development of language in children with autism spectrum disorder. *Autism Res.* **2015**, *9*, 993–1001. [CrossRef]
80. Leonard, H.C.; Bedford, R.; Pickles, A.; Hill, E.L. Predicting the rate of language development from early motor skills in at-risk infants who develop autism spectrum disorder. *Res. Autism Spectr. Disord.* **2015**, *13–14*, 15–24. [CrossRef]
81. MacDonald, M.; Lord, C.; Ulrich, D. The relationship of motor skills and adaptive behavior skills in young children with autism spectrum disorders. *Res. Autism Spectr. Disord.* **2013**, *7*, 1383–1390. [CrossRef]
82. Bradshaw, J.; Klaiman, C.; Gillespie, S.; Brane, N.; Lewis, M.; Saulnier, C. Walking Ability is Associated with Social Communication Skills in Infants at High Risk for Autism Spectrum Disorder. *Infancy* **2018**, *23*, 674–691. [CrossRef]
83. Ming, X.; Brimacombe, M.; Wagner, G.C. Prevalence of motor impairment in autism spectrum disorders. *Brain Dev.* **2007**, *29*, 565–570. [CrossRef] [PubMed]
84. Hilton, C.L.; Cumpata, K.; Klohr, C.; Gaetke, S.; Artner, A.; Johnson, H.; Dobbs, S. Effects of Exergaming on Executive Function and Motor Skills in Children With Autism Spectrum Disorder: A Pilot Study. *Am. J. Occup. Ther.* **2013**, *68*, 57–65. [CrossRef] [PubMed]

85. Fournier, K.A.; Hass, C.J.; Naik, S.K.; Lodha, N.; Cauraugh, J.H. Motor Coordination in Autism Spectrum Disorders: A Synthesis and Meta-Analysis. *J. Autism Dev. Disord.* **2010**, *40*, 1227–1240. [CrossRef] [PubMed]
86. Pierce, K.; Courchesne, E. Evidence for a cerebellar role in reduced exploration and stereotyped behavior in autism. *Biol. Psychiatry* **2001**, *49*, 655–664. [CrossRef]
87. Minshew, N.J.; Sung, K.; Jones, B.L.; Furman, J.M. Underdevelopment of the postural control system in autism. *Neurology* **2004**, *63*, 2056–2061. [CrossRef]
88. Rinehart, N.J.; Bradshaw, J.L.; Brereton, A.V.; Tonge, B.J. Movement Preparation in High-Functioning Autism and Asperger Disorder: A Serial Choice Reaction Time Task Involving Motor Reprogramming. *J. Autism Dev. Disord.* **2001**, *31*, 79–88. [CrossRef]
89. Cook, J.L.; Blakemore, S.J.; Press, C. Atypical basic movement kinematics in autism spectrum conditions. *Brain* **2013**, *136*, 2816–2824. [CrossRef]
90. Anzulewicz, A.; Sobota, K.; Delafield-Butt, J.T. Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. *Sci. Rep.* **2016**, *6*, 31107. [CrossRef]
91. Elison, J.T.; Wolff, J.J.; Reznick, J.S.; Botteron, K.N.; Estes, A.M.; Gu, H.; Hazlett, H.C.; Meadows, A.J.; Paterson, S.J.; Zwaigenbaum, L.; Piven, J. Repetitive Behavior in 12-Month-Olds Later Classified With Autism Spectrum Disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **2014**, *53*, 1216–1224. [CrossRef]
92. Petric, F.; Hrvatinic, K.; Babic, A.; Malovan, L.; Miklic, D.; Kovacic, Z.; Cepanec, M.; Stosic, J.; Simlesa, S. Four tasks of a robot-assisted autism spectrum disorder diagnostic protocol: First clinical tests. In Proceedings of the IEEE Global Humanitarian Technology Conference (GHTC 2014), San Francisco, CA, USA, 10–13 October 2014. [CrossRef]
93. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. In Proceedings of the ACM International Conference on Multimedia, Glasgow, UK, 1–4 April 2014. [CrossRef]
94. Michel, P.; El Kaliouby, R. Real time facial expression recognition in video using support vector machines. In Proceedings of the 5th international conference on Multimodal interfaces, Vancouver, BC, Canada, 5–7 November 2003; pp. 258–264.
95. Bartlett, M.S.; Littlewort, G.; Frank, M.; Lainscsek, C.; Fasel, I.; Movellan, J. Recognizing facial expression: machine learning and application to spontaneous behavior. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 568–573.
96. Bargal, S.A.; Barsoum, E.; Ferrer, C.C.; Zhang, C. Emotion recognition in the wild from videos using images. In Proceedings of the 18th ACM International Conference on Multimodal Interaction, Tokyo Japan, 12–16 November 2016; pp. 433–436.
97. Littlewort, G.; Bartlett, M.S.; Fasel, I.R.; Chenu, J.; Kanda, T.; Ishiguro, H.; Movellan, J.R. Towards social robots: Automatic evaluation of human-robot interaction by facial expression classification. *Adv. Neural Inf. Process. Syst.* **2004**, pp. 1563–1570.
98. Zhang, L.; Jiang, M.; Farid, D.; Hossain, M.A. Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. *Expert Syst. Appl.* **2013**, *40*, 5160–5168. [CrossRef]
99. Corneanu, C.A.; Simon, M.O.; Cohn, J.F.; Guerrero, S.E. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1548–1568. [CrossRef] [PubMed]
100. Nezami, O.M.; Dras, M.; Hamey, L.; Richards, D.; Wan, S.; Paris, C. Automatic Recognition of Student Engagement using Deep Learning and Facial Expression. *arXiv* **2018**, arXiv:1808.02324.
101. Liu, T.; Kappas, A. Predicting Engagement Breakdown in HRI Using Thin-slices of Facial Expressions. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
102. Eisenbarth, H.; Alpers, G.W. Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion* **2011**, *11*, 860–865. [CrossRef] [PubMed]
103. Åsberg Johnels, J.; Hovey, D.; Zürcher, N.; Hippolyte, L.; Lemonnier, E.; Gillberg, C.; Hadjikhani, N. Autism and emotional face-viewing. *Autism Res.* **2017**, *10*, 901–910. [CrossRef]
104. Ekman, P.; Friesen, W.V. Constants across cultures in the face and emotion. *J. Personal. Soc. Psychol.* **1971**, *17*, 124–129. [CrossRef]

105. Pantie, M.; Rothkrantz, L. Automatic analysis of facial expressions: the state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1424–1445. [CrossRef]
106. Zeng, Z.; Pantic, M.; Roisman, G.; Huang, T. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 39–58. [CrossRef]
107. Grynszpan, O.; Weiss, P.L.T.; Perez-Diaz, F.; Gal, E. Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. *Autism* **2013**, *18*, 346–361. [CrossRef]
108. Robe, A.; Dobrean, A.; Cristea, I.A.; Păsărelu, C.R.; Predescu, E. Attention-deficit/hyperactivity disorder and task-related heart rate variability: A systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **2019**, *99*, 11–22. [CrossRef]
109. Baxter, P.; Lemaignan, S.; Trafton, J.G. Cognitive Architectures for Social Human-Robot Interaction. In Proceedings of the Eleventh ACM/IEEE International Conference on Human Robot Interaction, Christchurch, New Zealand, 7–10 March 2016; pp. 579–580.