Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Project No. 611391

DREAM

Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Agreement Type: Collaborative Project
Agreement Number: 611391

D2.2.1 Tools for the assessment of child-robot interaction and diagnostics

Due Date: 01/04/2017

Submission date: **00/00/2017**

Start date of project: **01/04/2014** Duration: **54 months**

Organisation name of lead contractor for this deliverable: **Babes Bolyai University**

Responsible Person: **Daniel David**

Revision: **1.0**

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Contents

Executive Summary 3
Principal Contributors 4
Revision History... 5
Part 1 Diagnosis ......……………………………………………………………………...6
Part 2 Intervention ……………………………………………………………………....9
References…………………………………………………………………………….....12

Executive Summary

This deliverable summarizes our efforts in automatizing the tools that we used in the diagnosis phase and intervention sessions in our studies. In the first part of the deliverable (part A - diagnosis) we have provided a detailed description of the behavioural cues that clinicians use in the diagnosis process. These cues derive from the main symptoms described in DSM V. We then present an appraisal of technological means to augment this process, by identifying the appropriate technical ways to measure the behavioural cues. In the second part of the deliverable (part B – intervention) we have listed which are the behaviours that are needed to be captured during one intervention session with the robot and provided some preliminary results regarding the comparison between the therapist annotation and system's annotation. The final version of this deliverable will be provided in month 48.

Principal Contributors

The main authors of this deliverable are as follows (in alphabetical order)

Cristina Costescu, Babes-Bolyai University
Daniel David, Babes-Bolyai University
Anca Dobrean, Babes-Bolyai University
Silviu Matu, Babes-Bolyai University
Aurora Szentagotai, Babes-Bolyai University
Serge Thill, University of Skovde

Revision History

Version 1.0 (28-03-2017)
First draft, describing the theoretical background for the development of automated diagnosis and intervention tools.

Part A: Diagnosis

- **Evolution of assessment metrics**

There are several instrument that are based on DSM criteria used in ASD diagnosis and are considered to be the most relevant diagnosis instruments in the domain, such as Diagnosis Interview Revised (ADI-R)(Rutter & Le Couteur, 2003), the Childhood Autism Rating Scale (CARS) (Schopler, Reichler, DeVellis, & Daly, 1980), the Autism Diagnostic Observation Schedule Generic (ADOS-G) (Lord et al., 2000), the Diagnostic Interview for Social and Communication Disorders (DISCO) (Wing, Leekam, Libby, Gould, & Larcombe, 2002); the Gilliam Autism Rating Scale (GARS) (Gilliam, 1995). Although these behavioral observation are widely used among clinicians there are still discrepancies when it comes to diagnosis before the age of 3. For example, Klin, Lang, Chicchetti, and Volkmar (2000) state that when it comes to the agreement between clinicians in terms of DSM criteria it varies between 0.58 to 0.79. The situation is not better even when it comes to using the "golden standard" instrument for diagnosing ASD, for example ADOS. Lord, Rutter, P.C. Dilavore, and Risi (2002) show that inter-rater reliability is for some ratings as low as 0.38 in modules when younger children are assessed. Among the possible explanations for this discrepancies are the way data is collected, through simultaneous observation, coding and interpretation of the behaviors. Also all this data is collected from specific task that also vary to some extend in the way they are applied, depending on the clinicians' experience. Moreover, the process of learning how to code and interpret specific behaviors in order to obtain a high inter-rater reliability might last even years.

These observational judgments are then quantified according to standard protocols such as the Diagnosis Interview Revised (ADI-R) (Le Couteur, Lord, & Rutter, 2003), the Childhood Autism Rating Scale (CARS) (Schopler, Reichler, DeVellis, & Daly, 1980), the Autism Diagnostic Observation Schedule Generic (ADOS-G) (Catherine Lord et al., 2000), the Diagnostic Interview for Social and Communication Disorders (DISCO) (Wing, Leekam, Libby, Gould, & Larcombe, 2002), and the Gilliam Autism Rating Scale (GARS) (Gilliam, 1995).

Given the role of an accurate diagnosis of ASD for selecting appropriate treatment for individuals and the criticality of early interventions, it is crucial that the data collected be as valid as possible. Therefore, there is a need for methodologies that produce a quantified characterisation of the core symptoms in ASD during the diagnosis process. One way forward is to include machine-perception-guided technologies to augment the existing observational diagnoses and judgments made by clinicians.

One first step for using technological tools for augmenting the existing observational diagnoses is to detail the behavioural cues that clinicians use in the diagnosis process. These cues derive from the main symptoms described in DSM V, which underlie the development of the previously mentioned diagnostic instruments. We then present an appraisal of technological means to augment this process. While it is clear that technology cannot be expected to replace a clinicians' judgement, we show that the means to quantify several variables of interest do already exist.

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

	Required modalities									Classif.	
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 **Persistent deficits in social communication and social interaction across contexts**											**A**
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	▯	▯				▯					▯
Category **Restricted, repetitive patterns of behaviour, interests, or activities as manifested**											**B**
B1 Stereotyped or repetitive speech, motor movements, or use of objects											
1. Repetitive hand movements					▯						
2. Stereotyped or complex whole body movements				▯							

3. Repetitive vocalizations 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, ritualized 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				□	□		□				

Therefore in order to assess all the main symptoms for ASD, specifically: deficits in social-emotional reciprocity; deficits in nonverbal communicative behaviours used for social interaction; deficits in nonverbal communicative behaviours used for social interaction restricted, repetitive patterns of behaviour, interests, or activities; stereotyped or repetitive speech, motor movements, or use of objects; excessive adherence to routines, ritualized patterns of verbal or nonverbal behaviour, or excessive resistance to change; highly restricted, fixated interests that are abnormal in intensity or focus and hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of environment, as defined in DSM 5, we have to develop technological tools able to detect: gaze tracking, speech detection, speech analysis, posture tracking, gesture tracking, facial expressions, object tracking, and sound detection. All these tools were developed in WP4 and WP5 and will be used in the clinical studies.

Part B: Intervention (measuring performance in the intervention session)

One of the most important deficits in ASD is the social deficit. As reviewed by Bellini and Peters (2007) regarding to modalities to assess social skills, there are three important methods of measuring social functioning according to Gresham and their colleagues (2001). The first type of measure refers to the rating scales, questionnaires, and checklists that are completed by the individuals with ASD (self-report) or by their parents, teachers or caregivers. The advantage of using this type of measurement is that the information regarding their social performances is easy to obtain from multiple settings and from a variety of sources. However the disadvantage of this type of measurement is the fact that it often provide too general information about the improvement of social behaviors over time, and is therefore are not so useful in tracking progress during therapy. Furthermore, self-reports are less useful in studies involving children, since children often have difficulties in completing these scales. Also regarding the self-report measurements, when it comes to children often they might have some difficulties in completing the scales.

Another type of collecting information regarding someone's social abilities, according to Gresham et al. (2001) are assessments that use role-play scenarios and address questions regarding social-cognition. A common example of this type of measurement is the Sally Anne false belief task (Baron-Cohen, Leslie & Frith, 1985), through which professionals try to assess perspective taking by creating scenarios in which children are asked to infer the feelings and thoughts of others. Nonetheless, these type of measurements are considered to have the lowest psychometrical properties of out the three measurements categories, also there are studies showing that there wasn't established a clear relation between these type of measurements and rating scales (Otero, Schatz, Merrill & Bellini, 2015).

The third category of modalities of assessing the social skills includes direct assessment of their social skills and behaviors. This type of measurement were developed to objectively measure and describe developmental delays of social behaviors of children with ASD, information that is very important for diagnostic assessment, intervention planning and treatment monitoring. Moreover, using direct assessments are effective in evaluating the level of functioning at a certain time or during a type of treatment by observing the subject'shis behaviors in a naturalistic or semi-structured environment. One of the major problems with this type of measurements is that they their psychometric properties are debatable, also because little has been done in respect to increasing their psychometric soundness. The most common procedure in this sense is interrater reliability; an indicator that provides information about provides a way of quantifying the degree of agreement between two or more coders who make independent ratings about the features of a set of subjects (Hallgreen, 2012).

Additionally, the last type of measurement described above is extensively used in applied research studies using single-subject methodology to investigate the effectiveness of social skills interventions. Single – subject designs are the most used methods of evaluating social skill; in a recent review of social skills intervention studies for individuals with ASD, Matson et al. (2007) noted that more than 90 % of them adopted single-case designs.

Therefore the accuracy of direct assessment – observational data has implications for the trustworthiness of the assessments obtained. Given the potentially transformative nature of the interventions programs developed and the data used to guide recommended changes in practice, it is crucial for the data to be as valid as possible. Identifying and investing in methods of improving these mechanisms that reinforce accurate data collection is imperative for this objective.

Data capture and analysis is an important part of decision taking about whether a treatment is working or not. Using different types of technological tools can help increase the amount of data collected, make it easier to collect, and help professionals quickly scan through data to make better decisions (Kientz, Hayes, Westeyn, Starner, & Abowd, 2007).

In task 2.2 we have identified which are the behaviors that are needed to be captured during one intervention session with the robot or with the therapist. Therefore the performances in the intervention sessions were defined depending on the type of ability trained, as it follows: the child performance in the imitation task will be coded with score 1- if the child executes the requested movement correctly and with score 0- if the child doesn't execute the requested movement. Joint-attention performance is coded with score 1- if the child looks at the picture indicated by the interaction partner and with score 0- if the child doesn't look at the picture indicated. In the case of turn-taking, a score 1- will be given if the child waits his turn (doesn't move his hands above the touchscreen of the tablet when is the partner's turn) and score 0- will be given if the child doesn't wait his turn (he/she moves his hands above the tablet).

Among the performance in each task, we will also measure the engagement of the child in each intervention session. Engagement in the task is defined as eye contact, positive emotions and being present in front of the interactional partner. Therefore in order to detect this behaviors during the intervention session (performance and engagement) the following primitives need to be detected, as defined in D1.3: getChildperformance for imitation task: hand wave, hand covering eyes, hands on head, arms extended horizontally, straight horizontal hand gesture, straight vertical hand gesture. getChildperformance for joint attention task: look left/right (gaze tracking). getChildperformance for turn taking task: point left, point right and no movement. getChildEngagement for all three tasks: facial expressions (positive emotions), mutual gaze (for eye contact) and body position (in order to identify if the child is in front of the interactional partner).

All the above-mentioned behavioral cues were manually coded for each session of our developed experiments. In the following lines, we will present a comparison between the manual annotations of the behaviors during one-therapy sessions and perceptual features generated automatically using computer vision technology. While the suggested approach still requires a significant amount of improvements, it takes shorter time than manual annotations, thereby allowing annotation of larger quantities of video. Most importantly, the resulting annotations are of higher quality and more consistent than manual annotations.

In order to asses the degree in which technology can be used to assist therapists in diagnostic tasks and in therapy sessions, our colleagues from University of Skovde, developed in WP 5 one model (described in D 5.1) that was applied for eye contact variable:

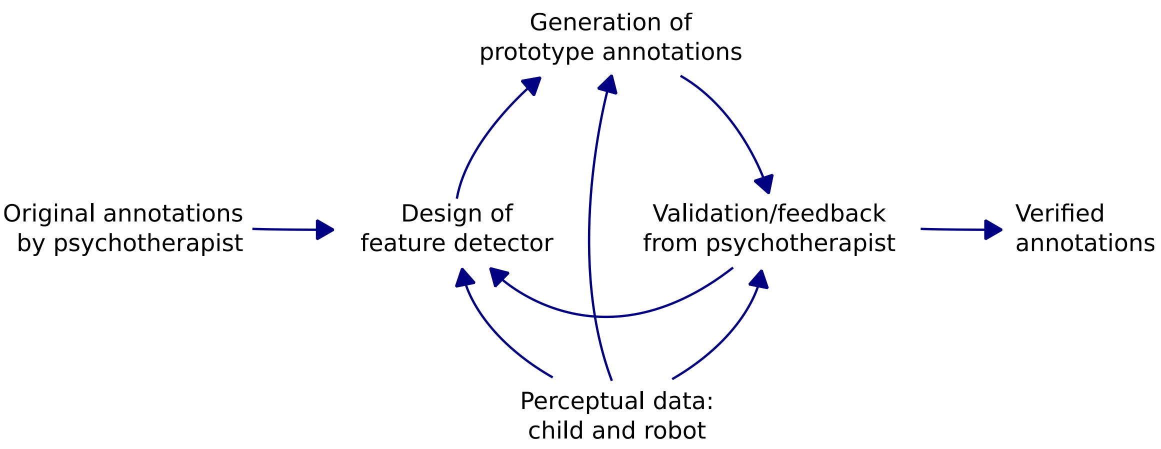

Figure 1: Illustration of the iterative process adopted for identification of features.

In order to compare the two the clinicians' feedback with the results generated by the system for one of our behaviour cues (eye contact) we have divided the feedback into six categories (defined in D 5.1):

- *Correct annotation* - including correctly identified start and end times within 0.5s for the child looking at, or glancing, at the robot or at the therapist;
- *Merge* indicates a single eye contact marked as two or more consecutive annotations;
- *Adjust* comprises annotations or an existing eye contacts, but where the start and/or end times are not correctly identified;
- *False positive* comprises incorrect annotations, not overlapping with an existing eye contact;
- *False negative* are existing eye contacts not identified by the classifier;
- *Ambiguous* comprises borderline-cases, where the clinicians were not able to determine whether there was an eye contact or not due to variation in the robot's positioning.

Some preliminary results are described in D 5.1. Therefore here we will provide only a brief description of the results emerged from the comparison of therapist annotation and system's annotation. In the first iteration of the algorithm, there were 31 eye contacts identified by the clinicians. Of these, twelve (26%) were correctly identified. A large proportion (39%) fell into Merge category. The second iteration comprised a larger set of 197 verified annotations. As a result of the updated detector, the number of annotations falling into category Merge is reduced to 4% while the number of correct annotations increased to 50%. More data are being processed for eye contact episodes and for other type of behaviours described above.

References

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?.*Cognition*,*21*(1), 37-46.

Bellini, S., Peters, J. K., Benner, L., & Hopf, A. (2007). A meta-analysis of school-based social skills interventions for children with autism spectrum disorders.*Remedial and Special Education*,*28*(3), 153-162.

Gilliam, J. E. (1995).*Gilliam autism rating scale: summary response form*. Pro-ed.

Gresham, F. M., Sugai, G., & Horner, R. H. (2001). Interpreting outcomes of social skills training for students with high-incidence disabilities.*Exceptional children*,*67*(3), 331-344.

Kientz, J. A., Hayes, G. R., Westeyn, T. L., Starner, T., & Abowd, G. D. (2007). Pervasive computing and autism: Assisting caregivers of children with special needs.*IEEE Pervasive Computing*,*6*(1).

Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial.*Journal of autism and Developmental disorders*,*30*(2), 163-167.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C.,& Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*,*30*(3), 205-223.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2002). Autism Diagnostic Observation Schedule (Western Psychological Services, Los Angeles).

Matson, J. L., Boisjoli, J. A., Gonzalez, M. L., Smith, K. R., & Wilkins, J. (2007). Norms and cut off scores for the autism spectrum disorders diagnosis for adults (ASD-DA) with intellectual disability.*Research in Autism Spectrum Disorders*,*1*(4), 330-338.

Otero, T. L., Schatz, R. B., Merrill, A. C., & Bellini, S. (2015). Social skills training for youth with autism spectrum disorders: a follow-up.*Child and adolescent psychiatric clinics of North America*,*24*(1), 99-115.

Rutter, M., Le Couteur, A., & Lord, C. (2003). Autism diagnostic interview-revised.*Los Angeles, CA: Western Psychological Services*,*29*, 30.

Sarac, S. B., Rasmussen, C. H., Rasmussen, M. A., Hallgreen, C. E., Søeborg, T., Colding-Jørgensen, M., & Mosekilde, E. (2012). A comprehensive approach to benefit–risk assessment in drug development.*Basic & clinical pharmacology & toxicology*,*111*(1), 65-72.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS).*Journal of autism and developmental disorders*,*10*(1), 91-103.

Wing, L., Leekam, S. R., Libby, S. J., Gould, J., & Larcombe, M. (2002). The diagnostic interview for social and communication disorders: Background, inter- rater reliability and clinical use.*Journal of Child Psychology and Psychiatry*,*43*(3), 307-325.