ENGAGEMENT DETECTION FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

Arodami Chorianopoulou[1], *Efthymios Tzinis*[2], *Elias Iosif*[2]
Asimenia Papoulidi[3], *Christina Papailiou*[4], *Alexandros Potamianos*[2]

[1]School of ECE, Technical University of Crete, Chania 73100, Greece
[2]School of ECE, National Tecnical University of Athens, Zografou 15780, Greece
[3]Dept. of Psychology, Panteion University, Athens 17671, Greece
[4]Dept. of Pre-school Education Sciences, University of the Aegean, Rhodes, 85132, Greece

achorianopoulou@isc.tuc.gr, etzinis@gmail.com, iosife@central.ntua.gr
a.papoulidi@panteion.gr, papailiou@rhodes.aegean.gr, potam@central.ntua.gr

ABSTRACT

Children with Autism Spectrum Disorder (ASD) face several difficulties in social communication. Hence, analyzing social interaction can provide insight on their social and cognitive skills. In this paper, we investigate the degree of engagement of children in interactions with their parents. Features derived from both participants including acoustic, linguistic and dialogue act features are explored. The effect of visual cues is also investigated. We experimented on the task of engagement detection using video-recorded sessions consisting of interactions of typically developing (TD) and ASD children. Results show that engagement is easier to predict for TD children than for ASD children, and that the parent's actions/movements are better predictors of the child's degree of engagement.

Index Terms: child engagement, engagement detection, autism spectrum disorder

1. INTRODUCTION

Engagement has a central role in the analysis of task-oriented social interactions, conveying information that can be connected with the behavioral and cognitive states of the participants. It can be defined as the process that involves two or more partners who jointly interact within a situational framework based on shared situation-specific aspects (e.g., perception of the environment, common goal, etc) [1]. The analysis of this process enables the better understanding of the underlying communicative mechanisms, which in turn can drive the design of relevant applications including multimedia analytics (e.g., speech analysis in human–human conversations [2]) and agents equipped with social skills (e.g., social robotics [3]). The automatic detection of engagement based on multimodal features extracted from audio-visual recordings of such interactions is a challenging task for children due to the unpredictability of their attentional patterns [4]. Individuals characterized by social communication impairments constitute an even more challenging user group regarding the success of such computational approaches.

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that disturbs the ability for social engagement, i.e., the development of interpersonal sympathy and collaborative action [5, 6, 7]. At birth infants express a simple interest in others' expressions, while by the age of two months become more sensitive to the reciprocity of emotions. Around three months, a typically developing (TD) infant often shifts attention to an object. At nine months, infants show a more pronounced interest in exploring specific emotional reactions and relating them to external targets [8, 9]. At this age an infant exhibits a new readiness to tune in with the intentions and interests of a partner in joint exploration and use of objects. Regarding the aforementioned aspects of social engagement, children with ASD demonstrate different degrees of deficits. Thus, identifying impairments in the ability to respond to social cues revealing different aspects of social engagement may allow the (early) distinction between young ASD and TD children [10, 11].

Features of speech prosody (rhythm, stress, and intonation) can be utilized as communication cues for identifying social engagement [12, 13]. It has been demonstrated that children with ASD exhibit more difficulty in perceiving some aspects of pragmatic/affective information compared to TD children [14, 15]. Verbal Response Latency (VRL) is another indicator of autism in children, defined in terms of response time in conversations. Long VRLs might occur when a complex conversation is performed [16] and they are related with the cognitive state of the conversational partners [17]. In [18], the degree of engagement for ASD children was investigated using acoustic features revealing a high correlation between vocal cues and engagement. In [19, 20], the ASD severity has been analyzed in relation to vocal arousal and emotion dynamics. Instructional settings have been regarded as appropriate fields for the study of engagement, which is a strong prerequisite for achieving several educational goals. For example, in [21, 22] facial and other automatically derived multimodal features were utilized for detecting the engagement level of students in such environments, while the detection results were found to be comparable to human observations. In [23], a platform was utilized for assisting ASD children to understand and express emotions.

This work extends previous approaches (e.g., [24]) that utilize acoustic and linguistic data for modeling the degree of social engagement in conversational interactions. Here, in addition to numerous other features, we utilize a set of social signals dealing with gaze and actions. Moreover, we adopt a psychologically-driven scheme for mapping communication intents to engagement levels. This scheme is applied over a newly created database of recorded sessions dealing with the interaction between TD and ASD children and their parents. We study the relative performance of the aforementioned features for the detection of child engagement in a scenario according to which the parent had to convince his/her child to play with a toy. A key assumption for the proposed approach is that the engagement of child can be triggered and regulated by his/her communicative partner (i.e., the parent). This assumption is investigated for

both TD and ASD children.

The remainder of the paper is organized as follows. The data collection and annotation process is described in Section 2, while the feature extraction is presented in Section 3. The experimental procedure along with the evaluation results are reported in Section 4. The conclusions of the present study are provided in Section 5.

2. DATASET

2.1. Video Recordings

A structured naturalistic procedure was decided to be the most appropriate method for video recordings [25]. Recordings took place in the child's home, and were characterized as structured because the introduction of certain situations by the psychologist did not leave the dyad complete freedom in play activities. The structured naturalistic method also ensured that all children would experience similar situations. Parents were asked to play with their child as they would normally do, by introducing a toy (car) which is provided Each session lasted approximately 45 minutes. A high quality video camera was used by an experienced psychologist so as to obtain high quality data for analysis.

On each video recording, researchers located the points on the footage where the parent uttered the word car and defined a framework around the parent's utterance called episode. An episode began when either the parent or the child first looked or acted at the car and ended when both the parent and the child shifted their attention from the car. The average duration of an episode was 4.86 minutes. The average duration of each episode for the ASD group was 5.99 minutes and for the TD group was 3.72 minutes. This difference was not statistically significant ($t = 1.11$, $p = 0.283$). Microanalysis within an episode consisted in noting the onset and offset of each manifested behavior from every category. This analysis provided information on when the parent's and child's attention converged on the car, the initiator and the responder of the interaction as well as the type of ongoing interaction (e.g. solitary play, converging interest or joint attention).

	ASD	TD	ALL
#utterances	966	645	1611
#sessions	33	33	66
#children	9	8	17
#male	8	6	14
#female	1	2	3

Table 1. Dataset description.

Table 1 presents the dataset's characteristics, namely, the number of parent's utterances, sessions and children. The age of the TD children ranged from 14 to 20 months, while the age of the ASD children was from 30 to 80 months. All children were at the single-word language development stage, and they were matched for visual-spatial and fine motor abilities. The language of the dataset is Greek.

2.2. Data Labeling

One expert annotator[1] labeled the dataset using the ELAN software [26] according to the following annotation types: 1) gaze at partner and/or object, 2) action on object, 3) action on partner, 4) emotion, and 5) transcription of utterances. The annotations were conducted for both partners, i.e., parent and child.

A subset of the aforementioned annotations (gaze, actions, and emotions) was manually associated (see Table 2) by psychologists with the following high-level categories of communication intent:

1. *Solitary*: behavior used to learn and explore the environment.
2. *Converging Interest*: two people express interest at the same object but they do not communicate between them about that.
3. *Regulatory*: behavior used to influence the behavior of others.
4. *Interpersonal*: reflects the motive for companionship (no objects included in the interaction).
5. *Interactional*: reflects the motive to achieve a goal in collaboration with another person.

The above categories are based on the seminal work of M. Halliday [27] aiming to encode the basic functions of language. These functions are defined with respect to a communicative environment where the development of child's language take place (e.g., interaction between parents and child). In addition, the experts assigned a discrete value (or range of values) of engagement to each of the aforementioned intent categories (see Table 2, where 0 denotes the absence of engagement, while 8 corresponds to the highest engagement degree).

For each recording, the aforementioned annotation types (i.e., gaze at object (Gaze), action of object (Obj.) and partner (Par.), emotion (Em.)) were considered for those excerpts where the parent was talking. The duration of an excerpt was determined by the start and the end of the respective parent's utterance, including N seconds after the end of the utterance (see Section 4.1).

For the creation of an evaluation dataset the following procedure was followed. Each excerpt (parent utterance) was associated with the child's intent and engagement degree according to the coding scheme presented in Table 2. This scheme applies to the child, i.e., the gaze and the actions of the child, as well as his/her emotional state were used as cues for determining his/her intent and engagement. For example, during an excerpt the child is assigned to engagement degree 1 if he/she gazes at the partner and/or object, and acts on the object. As it is shown in the same table, no excerpts were associated with the Regulatory category that corresponds to engagement degree 3. Also, 40 excerpts were assigned to two categories of intent (and engagement degrees) due to ambiguity reasons.

Intent	Engage	Gaze	Obj.	Par.	Em.	#utt.
Solitary	1	✓	✓			167
Converging	2	✓	✓			510
Regulatory	3	✓	✓	✓		0
Interpersonal	4	✓	✓	✓	✓	82
Interactional	5,6,7,8	✓	✓	✓		69
Not-engaged	0					784

Table 2. Data annotations and intention/engagement labeling. Obj: action on object, Par: action on partner, EM: emotion, #utt: number of utterances.

An example of parent utterances annotated with intent categories and the associated engagement degrees is shown in Table 3. In the second utterance, the parent says "The car, come.", while the child is holding/inspecting the specific object A screenshot from another example is depicted in Figure 1 along with a plot showing the child's degree of engagement as a function of time. The highest engagement

[1] Interrater reliability was assessed with video-recorded data from a subset of the sessions. Cohens kappa was 0.75 on average.

degree (equal to 7) occurs at 300 sec, which corresponds to the Interactional intent category. At this time point, the child looks at the parent and offers her the car.

Transcription	Gaze	Obj.	Par.	Intent	Engage
Do you want the car? What do you want?	LO LO LO	HI	MA	Conve-rging	3
The car, come.	LPE	HI OG		Intera-ctional	6

Table 3. Intention and engagement labeling examples; LO: looking at object, LPE: looking at partner's eyes/face, HI: holding/inspecting object, OG: offering/giving, MA: moving away.

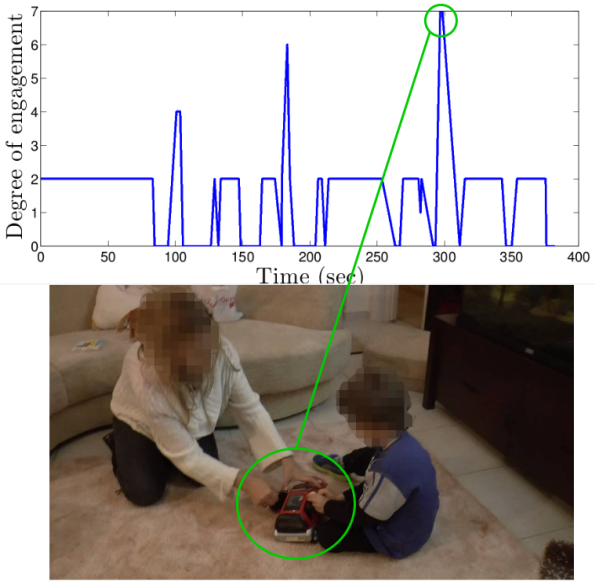

Fig. 1. Degree of engagement over time for one session. The example video frame corresponds to the highest engagement degree.

In order to evaluate the human perception with respect to child engagement two more annotators were employed. For a subset of the dataset, the annotators' task was to detect whether the child was engaged or not (i.e., binary decision) by 1) only hearing the parent's utterance (audio only), and 2) only watching the parent's movements (video only). The inter annotator's agreement was computed according to the Cohen's coefficient and it is presented in Table 4. The κ values indicate that the agreement for the audio-based detection is poor, i.e., the engagement can not be detected via audio only. The agreement regarding the video-based detection can be regarded as fair. These observations are helpful for the validation of the experimental results reported in Section 4.2.

3. FEATURE EXTRACTION

In this section, we briefly describe various feature sets that are used for the automatic detection of engagement. A synopsis of the features is presented in Table 5. Linguistic features were extracted for both parent and child, while the remaining feature sets, i.e., audio, video and affective text, were applied only on the parent's utterances.

		TD		ASD	
Modality	Task	Agree	κ	Agree	κ
Audio	detection	0.42	-0.24	0.51	-0.02
Video	detection	0.65	0.29	0.56	0.10

Table 4. Inter-annotator's agreement wrt. engagement detection.

Audio - Duration	
Acoustic	Energy, Pitch, Probability of Voicing, HNR, LPCs [1-10]
Duration	Utterance duration, VAD
Text	
Affective	Arousal, Valence, Dominance
Linguistic	#words, utterance repetition, #word repetition, #oov
Video	
Action-related	Gaze, action on object/partner

Table 5. List of features.

Audio and duration features

Acoustic: In order to model the style and quality of speech a set of frame-level features (low-level descriptors, LLDs) were extracted in a fixed window size of 30 ms with a 10 ms frame update, using the OpenSmile toolkit [28]. The proposed feature set contains the following LLDs: energy, pitch, probability of voicing, harmonics to noise ratio (HNR) and the first ten LPC coefficients. In order to extract utterance-level features, the following functionals were applied: extremes, moments and percentiles.

Duration: Children with ASD tend to respond after a longer period of time compared to TD children. Hence, a voice activity detection (VAD) feature either for the child (interpreted as response) or for the parent is employed. In both cases the feature is binary and activated only in the time window used for extracting the engagement labels. Additionally, the parent's speech duration is used.

Text features

Linguistic: Based on the assumption that speech is altered when speaking to children with ASD, we created a set of linguistic features using the transcribed utterances. These features include the number of words per utterance, a binary feature taking value 1 when the utterance is repeated and 0 otherwise, and the number of repeated words per utterance. Additionally, we observed that parents tend to use baby-talk (motherese) speech to describe sounds, for example the sound of a car. In order to recognize these words we compared our lexicon, consisting of $1,200$ words, with a Greek vocabulary of approximately $300,000$ words. Words that were not found in the vocabulary were annotated as out-of-vocabulary (OOV) and characterized as baby-speech.

Affective: The goal was to estimate the emotional content of the transcribed speaker utterances. A word w can be characterized regarding its affective content in a continuous space consisting of three dimensions, namely, valence, arousal, and dominance. In order to extract utterance-level ratings, the mean value of the ratings of the constituent words is computed, using an affective lexicon.

Details about the lexicon can be found in [29].

Action-related video features: As action-related features, we refer to the annotations regarding gaze and actions on objects/partner. Although these features were manually derived, they were included in the experimental features in order to investigate their role in engagement prediction. A detailed description of the annotations and their labels can be found in [25]. Information such as movements away or towards a person/object, gaze direction and symbolic or functional play are included.

4. EXPERIMENTS & EVALUATION

4.1. Experimental Procedure

The goal is to detect the child's engagement for each of the recording excerpts described in Section 2.2. The detection was considered as a binary classification problem, i.e., *engaged vs. non-engaged*. The excerpts annotated with non-zero degrees of engagement (see Table 2) were mapped to the *engaged* class, while the *non-engaged* class was assigned to the rest excerpts. Regarding the duration of the excerpts, the N parameter (i.e., the time that follows the end of the parents' utterances) was set to 1 sec.

For the experimental procedure, we adopted a leave-one-child-out scheme, i.e., testing over the data that refer to one child, while the training was performed using the data dealing with the rest children. For the experiments an SVM classifier with polynomial kernel from the Weka toolkit [30] is used. The classifiers were trained using the list of features presented in Table 5. Additionally, a forward selection algorithm was applied on the acoustic feature set. As fusion we used the concatenation of the different features sets.

4.2. Evaluation Results

The unweighted average classification accuracy *(UA)* and the unweighted average recall *(UR)* were used as evaluation metrics. The evaluation results for all features sets, as well as their fusion are presented in Table 6 for both parent's and child's features, Also, we report the baseline performance that corresponds to a majority class classifier that assigns each test sample to the majority class.

	UA (%)		UR	
	TD	**ASD**	**TD**	**ASD**
Majority class baseline	56.7	52.2	0.50	0.50
Parent's features				
Acoustic	47.6	47.1	0.46	0.50
Duration	56.6	46.8	0.44	0.47
Linguistic	**56.9**	50.7	**0.55**	**0.51**
Text Affective	50.4	46.3	0.49	0.50
Actions	**61.4**	**53.0**	**0.62**	**0.59**
Child's features				
Linguistic	49.2	44.7	**0.52**	0.48
Fusion				
All features	**63.3**	**53.9**	**0.64**	**0.57**

Table 6. Classification accuracy *(UA)* and unweighted recall *(UR)* results for the *engagement vs. no-engagement* task.

We observe that the best results (0.64 and 0.57 UR, and 63.3 and 53.9 *UA* for TD and ASD children, respectively) are achieved when fusing all feature types, which also exceed the majority class baseline. Regarding the individual feature types, the highest performance is observed for the action-related features yielding 0.62 and 0.59 *UR* for TD and ASD, respectively. This agrees with the validation study presented in Section 2.2, where the engagement was detected manually using audio- and video-based cues. The linguistic features, extracted from the parent's transcribed utterances, also achieve good performance (0.55 and 0.51 *UR* for TD and ASD, respectively).

Regarding the acoustic and duration feature sets, our expectation was that the parent's prosody could be a discriminative feature for the two engagement classes. However, this was not met. A manual data inspection revealed that the parents tend to speak motherese regardless of the child's degree of engagement. This is related to the fact that the the sessions were recorded between children and their parents instead of a psychologist. We believe that psychologists are inclined to use more strategic and less affective speech compared to parents.

Overall, the results indicate that the detection of engagement for TD children is more accurate compared to ASD children.

4.3. Feature Analysis

Also, we studied the relation between two basic features and the child's engagement. As basic features the following ones were used: presence of child's/parent's speech (VAD), and presence of repetitions in the parent's speech. The Pearson's correlation coefficient was computed between those features and child's engagement (see Table 7). Interestingly, the highest correlation is observed for the case of child's speech (Child VAD) despite the fact that the children who participated in this work were in the single-word language development stage. This holds for both TD and ASD children.

	Child VAD	Parent repetition	Parent VAD
TD	0.18	0.06	0.12
ASD	0.11	0.03	0.09

Table 7. Pearson correlation between the engagement labels and the VAD and repetition features.

5. CONCLUSIONS

We investigated the engagement of TD and ASD children in sessions with their parents and focused on the utterance-level engagement classification task. We used feature sets from different modalities, namely audio, text and video, extracted mostly from the parent's utterances rather than the child's. Our results suggest that the child's engagement can be predicted by analyzing the parent's behavior only with moderate accuracy. Prediction accuracy was higher in TD than in ASD children Video-related and lexical features from the parent's transcribed utterances were the most informative, while acoustic features performed poorly. We expected children's speech to be more correlated with the child's engagement level, however most of the children used non-vocal cues or reacted to the task's needs with movements and gazing. In future work, more features will be investigated and alternative machine learning algorithms will be evaluated for engagement prediction. The action-related features as well as the transcribed utterances will be automatically extracted using machine learning algorithms.

Acknowledgements. This work has been partially supported by the BabyRobot project supported by the EU Horizon 2020 Programme with grant #687831.

6. REFERENCES

[1] Candace L. Sidner, Christopher Lee, and Neal Lesh, "The role of dialog in human robot interaction," in *International workshop on language understanding and agents for real world interaction*, 2003.

[2] Matthew P. Black, Athanasios Katsamanis, Brian R. Baucom, Chi-Chun Lee, Adam C. Lammert, Andrew Christensen, Panayiotis G. Georgiou, and Shrikanth Narayanan, "Toward automating a human behavioral coding system for married couples interactions using speech acoustic features," *Speech Communication*, vol. 55, no. 1, pp. 1–21, 2013.

[3] Alexandros Potamianos, Costas Tzafestas, Elias Iosif, Franziska Kirstein, Petros Maragos, Kerstin Dauthenhahn, Joakim Gustafson, John Erland stergaard, Stefan Kopp, Preben Wik, Oliver Pietquin, and Samer Al Moubayed, "Babyrobot - next generation social robots: Enhancing communication and collaboration development of TD and ASD children by developing and commercially exploiting the next generation of human-robot interaction technologies," in *Proceedings of the 2nd Workshop on Evaluating Child-Robot Interaction (CRI) at Human-Robot Interaction (HRI'16)*, 2016.

[4] Judith S. Nijmeijer, Ruud B. Minderaa, Jan K. Buitelaar, Aisling Mulligan, Catharina A. Hartman, and Pieter J. Hoekstra, "Attention-deficit/hyperactivity disorder and social dysfunctioning," *Clinical Psychology Review*, vol. 28, no. 4, pp. 692–708, 2008.

[5] Leo Kanner, "Autistic disturbances of affective contact," 1943.

[6] Uta Frith, "Autism and asperger syndrome," *Cambridge University Press*, 1991.

[7] Colwyn Trevarthen, "Autism as a neurodevelopmental disorder affecting communication and learning in early childhood: prenatal origins, post-natal course and effective educational support," *Prostaglandins, Leukotrienes and Essential Fatty Acids*, vol. 63, no. 1, pp. 41–46, 2000.

[8] Vasudevi Reddy and Martyn Barrett, "Prelinguistic communication," *The development of language*, pp. 25–50, 1999.

[9] Colwyn Trevarthen, "Infant semiosis," *Origins of semiosis: Sign evolution in nature and culture*, vol. 116, pp. 219, 1994.

[10] Geraldine Dawson, Karen Toth, Robert Abbott, Julie Osterling, Jeff Munson, Annette Estes, and Jane Liaw, "Early social attention impairments in autism: social orienting, joint attention, and attention to distress," *Developmental psychology*, vol. 40, no. 2, pp. 271, 2004.

[11] Peter C. Mundy and C. Françoise Acra, "Joint attention, social engagement, and the development of social competence," *The development of social engagement: Neurobiological perspectives*, pp. 81–117, 2006.

[12] Daniel Bone, Matthew P. Black, Chi-Chun Lee, Marian E. Williams, Pat Levitt, Sungbok Lee, and Shrikanth Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in *INTERSPEECH*, 2012, pp. 1043–1046.

[13] Daniel Bone, Matthew P. Black, Anil Ramakrishna, Ruth Grossman, and Shrikanth Narayanan, "Acoustic-prosodic correlates of awkward prosody in story retellings from adolescents with autism," 2015.

[14] Rhea Paul, Lawrence D. Shriberg, Jane McSweeny, Domenic Cicchetti, Ami Klin, and Fred Volkmar, "Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders," *Journal of Autism and Developmental Disorders*, vol. 35, no. 6, pp. 861–869, 2005.

[15] Susan Peppé, Joanne McCann, Fiona Gibbon, Anne OHare, and Marion Rutherford, "Receptive and expressive prosodic ability in children with high-functioning autism," *Journal of Speech, Language, and Hearing Research*, vol. 50, no. 4, pp. 1015–1028, 2007.

[16] Theodora Chaspari, Daniel Bone, James Gibson, Chi-Chun Lee, and Shrikanth Narayanan, "Using physiology and language cues for modeling verbal response latencies of children with ASD," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 3702–3706.

[17] Susan Ervin-Tripp, "Children's verbal turn-taking," *Developmental pragmatics*, pp. 391–414, 1979.

[18] Rahul Gupta, Chi-Chun Lee, Daniel Bone, Agata Rozga, Sungbok Lee, and Shrikanth Narayanan, "Acoustical analysis of engagement behavior in children.," in *Workshop on Child Computer Interaction (WOCCI)*, 2012, pp. 25–31.

[19] Erik Marchi, Björn Schuller, Anton Batliner, Shimrit Fridenzon, Shahar Tal, and Ofer Golan, "Emotion in the speech of children with autism spectrum conditions: prosody and everything else," in *Workshop on Child Computer Interaction (WOCCI)*, 2012, pp. 17–24.

[20] Daniel Bone, Chi-Chun Lee, Alexandros Potamianos, and Shrikanth Narayanan, "An investigation of vocal arousal dynamics in child-psychologist interactions using synchrony measures and a conversation-based model," in *INTERSPEECH*, 2014, pp. 218–222.

[21] Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan, "The faces of engagement: Automatic recognition of student engagementfrom facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.

[22] Dominique Vaufreydaz, Wafa Johal, and Claudine Combe, "Starting engagement detection towards a companion robot using multimodal features," *Robotics and Autonomous Systems*, vol. 75, pp. 4–16, 2016.

[23] Björn Schuller, Erik March, Simon Baron-Cohen, Amandine Lassalle, Helen OReilly, Delia Pigat, Peter Robinson, Ian Davies, Tadas Baltrusaitis, and Marwa Mahmoud, "Recent developments and results of ASC-inclusion: An integrated internet-based environment for social inclusion of children with autism spectrum conditions," in *Proceedings of the 3rd International Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI)*, 2015.

[24] Daniel Bone, Chi-Chun Lee, Theodora Chaspari, Matthew P. Black, Marian E. Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan, "Acoustic-prosodic, turn-taking, and language cues in child-psychologist interactions for varying social demand," in *INTERSPEECH*, 2013, pp. 2400–2404.

[25] "Deliverable WP3: Report on affective and cognitive modeling of TD and ASD children (months 1-9)," *BabyAffect Project*, 2015.

[26] Han Sloetjes and Peter Wittenburg, "Annotation by category ELAN and ISO DCR," in *6th International Conference on Language Resources and Evaluation (LREC)*, 2008.

[27] Michael Alexander Kirkwood Halliday, "Learning how to mean–explorations in the development of language," 1975.

[28] Florian Eyben, Martin Wöllmer, and Björn Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[29] Elisavet Palogiannidi, Polychronis Koutsakis, Elias Iosif, and Alexandros Potamianos, "Affective lexicon creation for the Greek language," in *10th International Conference on Language Resources and Evaluation (LREC)*, 2016.

[30] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.