

Voice Emotion Games: Language and Emotion in the Voice of Children with Autism Spectrum Condition

**Erik Marchi,
Björn Schuller**
MISP group
Technische Universität
München
Germany
erik.marchi@tum.de
schuller@tum.de

**Simon Baron-Cohen,
Amandine Lassalle,
Helen O'Reilly,
Delia Pigat**
Autism Research Centre
University of Cambridge
Cambridge, UK
sb205@cam.ac.uk

**Ofer Golan,
Shimrit Fridenson-Hayo,
Shahar Tal**
Department of Psychology
Bar-Ilan University
Ramat-Gan, Israel
ofer.golan@biu.ac.il

ABSTRACT

Three decades of research have shown that children and adults with Autism Spectrum Conditions (ASC) may experience significant difficulties recognizing and expressing emotions and mental states. These difficulties are apparent when individuals with ASC attempt to recognize emotions from facial expressions, from vocal intonation, from gestures and body language. This ASC-Inclusion project set up advanced ICT-enabled solutions and serious games that assist children with ASC to improve their socio-emotional communication skills, combining voice, face, and body gesture analysis, and giving corrective feedback regarding the appropriateness of the child's expressions. The present contribution focuses on the recognition of emotion in speech. For this purpose, a database of prompted phrases was collected in English, Swedish, and Hebrew, inducing nine emotions embedded in short-stories. It contains speech of children with ASC and typically developing children under the same conditions. We evaluate the emotion task over the nine categories, by investigating the discrimination of each emotion against the remaining ones. The results show performances up to 83.8% unweighted average recall.

Author Keywords

Autism Spectrum Conditions, inclusion, computerised environment, emotion recognition

INTRODUCTION

Over the last thirty years of research, it has been shown that children and adults with Autism Spectrum Conditions (ASC) may experience significant difficulties in recognising and expressing emotions from facial expressions, speech, gestures, and body language. Attempts to teach emotion and mental state recognition, either on an individual basis or as a part of social skills group training, have shown mixed results. A solution for the shortage of trained therapists for individuals

with ASC may be found in Information and Communication Technology (ICT), which enables users everywhere to enjoy state-of-the-art professional support on-line. The computerised environment is especially appealing for individuals with ASC, due to its predictable, controllable and structured nature, which facilitates them to use their strong systemizing skills. Existing systems, such as the Rachel Embodied Conversational Agent (ECA) [11] and the Mind-Reading software [3], aim to elicit the targeted emotion through an interactive agent in order to study the interaction patterns of children with ASC and to teach people in the spectrum to recognise complex emotions using interactive multimedia. The ASC-Inclusion project [14, 13] created an internet-based platform that assists children with ASC to improve their socio-emotional communication skills. Unlike the past ICT solutions, the project addresses the recognition and the expression of socio-emotional cues, by providing an interactive-game that gives scores on the prototypicality and on the naturalness of child's expressions. It combines several state-of-the-art technologies in one comprehensive virtual environment (VE), combining voice, face and body gesture analysis, giving corrective feedback as for the appropriateness of the child's expressions. In 2007, Potamianos and Narayanan [12] have examined major differences in children versus adult voices showing how acoustic, lexical and linguistic characteristics of solicited and spontaneous children's speech are correlated with age and gender. These differences are, however, even greater when looking at the population of children affected by ASC. This makes the automatic speech and emotion processing tasks much more complex when dealing with the ASC population. When one adds in the background noise of children's homes and doctors offices, it makes it even harder for automatic recognition systems to perform accurately. In the past two decades, a number of affective games encouraging users to express emotional states were developed. In "Emote to Win" the user influences the behaviour of a virtual pet by appropriate emotional speech reactions [7]. In [5], the game "XQuest" was used to induce natural emotional speech for acoustic analysis. A new emotionally responsive game was developed based on "Half-Life", enabling the recognition of affective cues from the speech of players [6]. Further studies focused on the detection of frustration, politeness, and neutral behaviours in a conversational computer game [18]. However, as is the case with other games, this study refers to the analysis of typical behaviour. Recent

studies also investigated how prosodic speech cues of children with ASC can be quantified during spontaneous interaction [1]. Yet, in spite of these challenges, we do, however, consider the integration of voice-based technologies in ICT-based virtual platform as an important component to enable multimodal interaction between children with ASC and VE. The ability to convey socio-affective behaviours through speech is probably the most natural way to engage social interactions [9], in addition to facial expressions and body gestures.

The present study focuses on the recognition of emotional expressions in the voice of children with ASC, in order to investigate the classification performances with large sets of features that include a vast number of acoustic, spectral and cepstral features. Given our interest in the classification of children's emotional vocal expression, a database of prompted phrases was collected in English, Swedish, and Hebrew, inducing up to nineteen emotions embedded in short-stories. The utterances were produced by children with Autism Spectrum Conditions as well as by typically developing children.

The article is structured as follows: first, a detailed description of the database is given (Section ASC-Inclusion children's emotional speech database); then we define the experimental tasks, features and set-up (Section Experiments). We next comment on the evaluation results (Section Results) before concluding the paper in Section Conclusions.

ASC-INCLUSION CHILDREN'S EMOTIONAL SPEECH DATABASE

As an evaluation database for the recognition of emotions and for the analysis of speech features that are modulated by emotion, a set of prototypical emotional utterances containing sentences spoken in English, Swedish, and Hebrew by children with ASC and typically developing children has been created. In this section we provide a description of the three language-dependent subsets.

Hebrew dataset

The Hebrew dataset was already introduced in [8, 10], and consists of seven children (6 male and 1 female) at the age of 6 to 10 ($M=8.1$, $SD=1.6$), all diagnosed with an autism spectrum condition by trained clinicians. 10 typically developing children (5 female and 5 male) at the age of 5 to 9 ($M=7.2$, $SD=1.8$) were selected to form the control group. In order to limit the effort of the children, the experimental task was designed to focus on the six "basic" emotions except *disgust*: *happy*, *sad*, *angry*, *surprised*, *afraid* plus other three mental states: *ashamed*, *calm*, *proud*, and *neutral*. During a 2 hour meeting with the child and his/her parents, a semi-structured observation was conducted which included free-play in a virtual environment, followed by a directed play in pre-selected games, and by an interview with the child. Only then, the recording session was held, since it requires a good rapport with the child. The recordings took place at the children's home according to the following set-up: the child and the examiner sat at a table in front of a laptop. The microphone stood next to the laptop, about 20 cm in front of the child. As recording device, a Zoom H1 Handy Recorder was used.

Recordings were taken in wav format at a sampling rate of 96 kHz and a quantization of 16 bits and stored directly on the microphone's internal SD memory card. The examiner read to the child a sequence of short stories from a power point presentation. The stories were simple and short. The child was asked to imagine that he/she was the main character in the story. The stories contained, every few sentences, a quotation of an utterance by the story's main character. Each of these quotations related to a specific emotion, which was explicitly stated. For example: [Danny said happily: "*It was the best birthday I ever had!*"] or [Jain was very surprised. She looked at the box and said: "*What is that thing?*"]. When the examiner read the stories, he read the sentence on a flat, unnatural tone. Then he asked the child to say the sentence as the child in the story would have said it. Each slide that contained an emotional utterance to be said by the child also showed a photograph of a person expressing the same emotion through his facial expressions. The photos were taken from the Mind-Reading database [3]. The text material used for the task consists of nine stories. Each story aims to elicit some of the target emotions as described above and contains from 3 to 7 different emotional utterances. In total, the nine stories contain 37 utterances.

An example for one of the nine stories is:

Happy - Today it's a special day for Danny: it's his birthday! Danny was very happy - a birthday is an especially enjoyable and fun day. Danny went into his sister's room and said **happily**: "*Today's my birthday!*".

Sad - Afterwards he entered the kitchen. He noticed his mother was preparing a simple breakfast for him and a not a birthday's one. Danny was very sad. He was convinced his family had forgotten his birthday. In school no one had congratulated him either, not even his teacher! Tears flooded his eyes, and so he looked for his sister on break time. When he found her, he told her **sadly**: "*No one had remembered*".

Angry - On his way home the sad feeling had faded away, and anger burned inside of him. He was so angry of his mom and classmates, and said **angrily** to his sister: "*I won't remember their birthday either!*".

Surprised - When he got back home, there was a complete silence. He went into the dark kitchen, lit up the light and suddenly heard: "surprise"! He saw there his parents and classmates holding balloons! He was very **surprised** - and said: "*What's going on?*".

Happy - Danny was happy, they haven't forgotten him, they planned him a surprise birthday party. After a party, he went to his sister and said **happily**, "*It was the best birthday I ever had!*".

Since the recordings were held at the children's home, they are partly affected by background noise. The Hebrew dataset comprises 529 utterances with a total duration of 16 min 24 sec and an average utterance length of 1.8 sec. 178 utterances contain emotional speech of children with ASC with a total recording time of 7 min 1 sec and an average utterance duration

Table 1: Number of utterances per emotion category (# Emotion), and overall number of utterances (# All) for the three languages. Emotion classes: happy (Ha), sad (Sa), angry (An), surprised (Su), afraid (Af), proud (Pr), ashamed (As), calm (Ca), neutral (Ne). Diagnosis categories: Typically developing children (TD), and children with Autism Spectrum Condition (ASC). Gender: number of male (m) and female (f) subjects.

# utterances	Diagnosis	# Subjects		# Emotion									# All
		m	f	Af	An	Ha	Sa	Su	As	Pr	Ca	Ne	
English	TD	5	4	40	40	50	40	40	40	50	28	40	368
-	ASC	5	3	31	30	39	31	31	31	39	24	32	288
Swedish	TD	6	5	40	40	49	48	39	28	48	30	38	360
-	ASC	9	-	36	36	45	44	36	27	45	27	36	332
Hebrew	TD	5	5	38	38	49	38	38	37	46	27	40	351
-	ASC	6	1	18	20	30	21	21	17	22	13	16	178

of 2.37 sec. The remaining 351 utterances are produced by the control group with a total duration of 9 min 23 sec and an average utterance recording time of 1.61 sec. The first two rows in Table 1 provide the number of utterances for each group and per emotion class.

Swedish dataset

A total number of 20 children took part in the recordings held in Sweden. The language throughout recordings is Swedish and all children are native speakers. The focus group consists of 9 children (9 male pupils) at the age of 7 to 11 ($M=9.1$, $SD=1.2$). All children were diagnosed with an autism spectrum condition by trained clinicians, based on established criteria (DSM IV/ICD 10). The control group consists of 11 children (5 male and 6 female pupils) at the age of 5 to 9 ($M=6.8$, $SD=1.7$). The recording protocol adopted for the dataset is identical to the one used in the Hebrew dataset. The experimental task focused again on the six “basic” emotions except *disgust*: *happy*, *sad*, *angry*, *surprised*, *afraid* plus other three mental states: *ashamed*, *calm*, *proud*, and *neutral*. As recording device, a Zoom H4 with RØDE NTG-2 microphone was used. Recordings were taken in wav format at a sampling rate of 96 kHz and a quantization of 16 bits. The Swedish dataset comprises a total of 692 utterances containing emotional speech of children with ASC and typically developing children. 360 utterances are related to material collected from children with ASC, whereas the remaining 332 utterances contains typical speech from children. The third and fourth rows of Table 1 show the number of utterances for each group (typically developing children (TD) and children with ASC (ASC)), and for each emotion.

English dataset

A total number of 18 children took part in the recordings held in England. The recordings are in English and all children are native speakers. The focus group consists of 8 children (5 male and 4 female pupils) at the age of 7 to 11 ($M=8.8$, $SD=1.5$). All children were diagnosed with an autism spectrum condition by trained clinicians, based on established criteria (DSM IV/ICD 10). The control group consists of 10 children (5 male and 5 female pupils) at the age of 5 to 10 ($M=7.9$, $SD=1.6$). The

recording protocol adopted for the dataset is identical to the one used in the Hebrew dataset, however the experimental task focused also on “individual” and “social” emotions: *excited*, *interested*, *bored*, *worried*, *disappointed*, *frustrated*, *hurt*, *kind*, *jealous*, *unfriendly*, and *joking* plus *disgust*. As recording device, a Zoom H1 Handy Recorder was used. Recordings were taken in wav format at a sampling rate of 96 kHz and a quantization of 16 bits. The English dataset comprises a total of 1507 utterances contain emotional speech of children with ASC and typically developing children. 660 utterances contain material collected from children with ASC, whereas the remaining 847 utterances contain typical speech from children. In this dataset 21 emotions were collected, however in order to compare performances across different datasets we only selected those commonly shared among the datasets. The last two rows of Table 1 shows the number of utterances for each group (typically developing children (TD) and children with ASC (ASC)) and for each emotion.

Compared to the standards of present day databases used for automatic speech processing, this is a small database; however, taking into account the difficulties to recruit children from the envisaged population, to successfully conduct all the experimental tasks, and in comparison to other studies within the fields of ASC and emotion modelling for specific and less-studied populations, it can be taken as fairly representative, especially considering it contains recordings in three languages.

EXPERIMENTS

In this part we describe the classification tasks in Section Tasks, the feature sets in Section Features, and the experimental setup (Section Setup).

Tasks

Eight tasks were evaluated, and they consist in the classification of every “*e* vs. *Rest*”, where a given emotion is evaluated against the remaining emotions. We choose only the emotions that are present in all the three dataset except calm that resulted to be acoustically similar to neutral, and showed poor classification performances. All the tasks were performed on

the focus and control group subsets separately; a detailed description of the number of instances belonging to the classes of each task per subset is given in Table 1.

Features

The experiments were conducted using the INTERSPEECH 2013 ComParE Challenge [16] feature set. The feature set is based on the acoustic feature set used for the INTERSPEECH 2012 Speaker Trait Challenge [15]. We used the open-source openSMILE feature extractor [2] and extracted the features on a per-chunk level. The feature set includes energy, spectral, cepstral (MFCC) and voicing related low-level descriptors (LLDs) as well as a few LLDs including logarithmic harmonic-to-noise ratio (HNR), spectral harmonicity, and psychoacoustic spectral sharpness. Altogether, the 2013 COMPARE feature set contains 6373 features. Table 2 provides a detailed list of LLDs used in the features set.

Table 2: List of low-level descriptors (LLD) included in the feature set.

4 energy related LLD
Sum of auditory spectrum (loudness).
Sum of RASTA-style filtered auditory spectrum.
RMS Energy.
Zero-Crossing Rate.
55 spectral LLD
RASTA-style auditory spectrum, bands 1-26 (0–8 kHz).
MFCC 1–14.
Spectral energy 250–650 Hz, 1 k–4 kHz.
Spectral Roll Off Point 0.25, 0.50, 0.75, 0.90.
Spectral Flux, Centroid, Entropy, Slope
Psychoacoustic Sharpness, Harmonicity,
Variance, Skewness, Kurtosis.
6 voicing related LLD
F0 by SHS + Viterbi smoothing, Probability of voicing
logarithmic HNR, Jitter (local, delta), Shimmer (local).

Setup

Since all data sets are unbalanced (i.e. one class is under-represented in the data), the unweighted average recall (UAR) of the classes is used as scoring metric. Adopting the Weka toolkit [4], Support Vector Machines (SVMs) with linear kernel were trained with the Sequential Minimal Optimization (SMO) algorithm. SVMs have been chosen as classifier since they are a well known standard method for emotion recognition due to their capability to handle high and low dimensional data. The SVM training has been made at different complexity constant values $C \in \{0.001, 0.002, 0.005, 0.01, 0.02\}$. To ensure speaker independent evaluations, Leave-One-Speaker-Out (LOSO) cross-validation has been performed. In each fold we up-sampled the training material in order to have a balanced set. Furthermore, we adopt the speaker z -normalisation (SN) method since it is known to improve the performance of speech-related recognition tasks, as described in [17]. With such a method, the feature values are normalised to a mean of zero and a standard deviation of one for each speaker.

RESULTS

In this section we comment the results obtained on the three language-dependent datasets. Figure 1 shows the best results obtained over the different complexities for each dataset, namely English, Swedish, and Hebrew. We perform the 2-class tasks “ e vs. $Rest$ ”, with $e \in E$, where $E = \{Afraid, Angry, Happy, Sad, Surprised, Ashamed, Proud, Neutral\}$ and $Rest = E - e$. All the tasks were performed both on the focus and on the control set separately.

English

Figure 1a shows detailed results obtained on the English dataset. On the control group, we observe performances over 70% UAR on Angry, Happy, Sad, Surprised, and Neutral. In particular Angry and Surprised show higher performances up to 78.9% and 79.3% UAR, respectively. Neutral seems to be highly discriminable from the remaining emotions resulting in 83.8% UAR. The other emotions, namely Afraid, Ashamed, and Proud show slightly lower performances down to 64.7% UAR. On the focus group, higher UAR is achieved on Afraid, Happy, and Ashamed up to 71.3%, 72.9%, and 80.2% UAR, respectively. Lower performances are observed for Sad and Neutral down to 65.0% UAR. In particular, Angry seems to be a difficult task with the lowest performance of 54.7% UAR. Comparing control and focus group results on the English dataset, we observe that the most difficult tasks are Angry, Sad, and Neutral where performances are clearly below the ones obtained on the data collected from typically developing children. From pure classification point of view, the other tasks seem to perform similarly. Overall, a delta of 5.5% absolute UAR applies to the classification performances averaged over all the tasks within the two groups. In particular we can observe an average UAR of 74.3% and 68.8% for the control and focus group, respectively.

Swedish

In comparison with the English dataset, we observe – on average – slightly lower performances on both groups. On the control group, we achieved over 65% UAR on Angry, Afraid, Happy, Sad, Ashamed, and Neutral. In particular, Happy and Sad show higher performances up to 70.6% and 70.7% UAR, respectively. Additionally, Neutral seems to be – as found in the English dataset as well – highly discriminable from the remaining emotion leading to 81.5% UAR. The other emotions, namely Surprised, and Proud show slightly lower performances down to 63.3% UAR. On the focus group, higher UAR is achieved on Afraid, Happy, Sad, and Ashamed up to 69.5%, 65.4%, 72.7%, and 69.5% UAR, respectively. Neutral shows the highest performance (77.9% UAR) among all the emotions in the focus group. Lower performances are observed for Surprised and Proud down to 60.6% UAR. As in the English dataset, Angry seems to be again a difficult task with the lowest performance of 56.0% UAR. Figure 1b shows detailed results of the evaluation performed on the Swedish dataset. Results on the control and focus group show similar performances; In fact, we can observe a delta of 1.8% UAR between classification performances averaged over all tasks within the control group (68.6% UAR) and focus group (66.8% UAR). An exception has to be made for Angry (54.7% UAR)

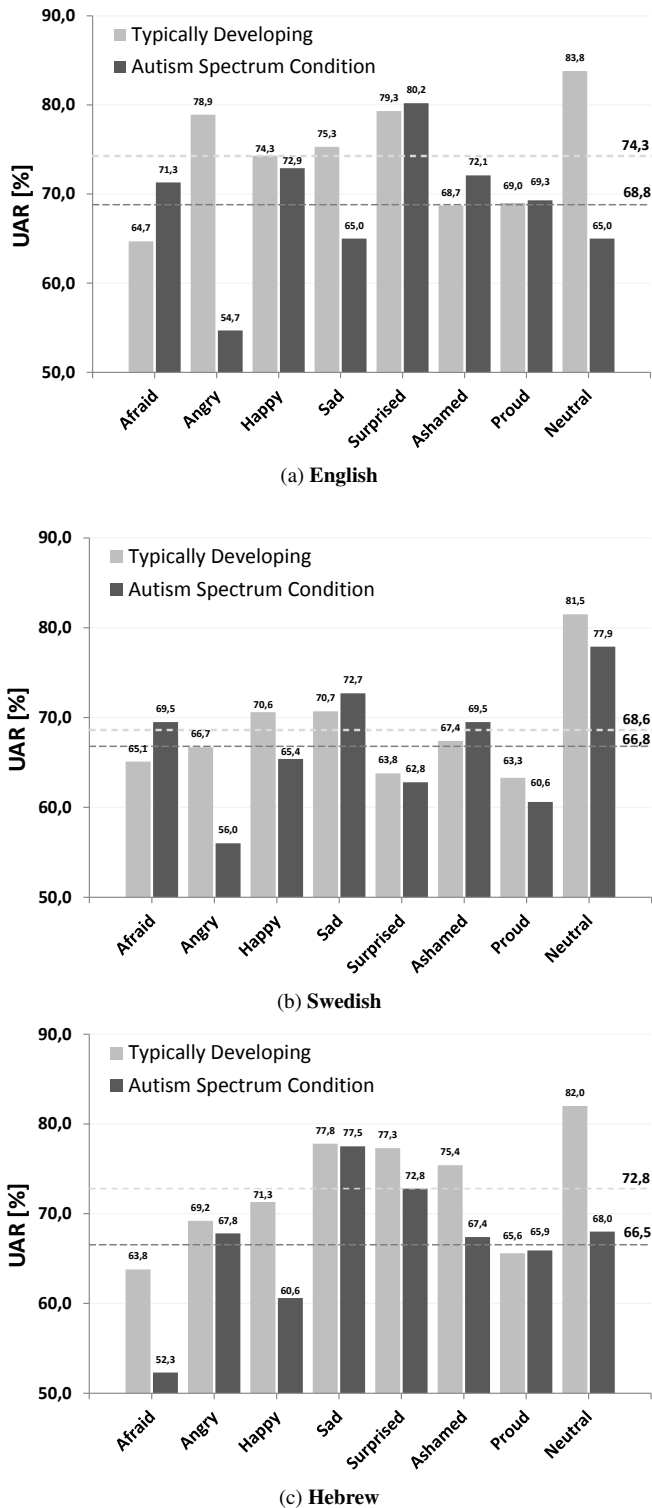


Figure 1: Classification result for binary “*e* vs. *Rest*” on the (a) English, (b) Swedish, and (c) Hebrew data sets. *Rest* class includes Neutral. For example “*happy* vs. *Rest*”: *Rest* = {*Angry*, *Afraid*, *Sad*, *Surprised*, *Ashamed*, *Proud*, *Neutral*}. Results are given for typically developing children and children with ASC. Dotted line indicate average UAR over all the tasks per group.

which is clearly below the ones obtained on the data collected from typically developing children (78.9% UAR).

Hebrew

On the control group, Happy, Sad, Surprised, Ashamed, and Neutral show performances greater than 70% UAR. Happy and Sad show higher performances up to 71.3% and 77.8% UAR, respectively (cf. Figure 1c). Also in this dataset, Neutral is highly discriminable from the remaining emotion with 82.0% UAR. The other emotions, namely Afraid, and Angry show slightly lower performances down to 63.8% UAR. On the focus group, higher UAR is achieved on Sad, and Surprised up to 77.5%, and 72.8% UAR, respectively. Lower performances are observed for Happy and Neutral down to 60.6% UAR. Afraid seems to be in this dataset the most difficult task with the lowest performance of 52.3% UAR. Results on the control and focus group show rather different performances: we can observe a delta of 6.3% UAR between classification performances averaged over all tasks within the control group (72.8% UAR) and focus group (66.5% UAR). Afraid, Happy, and Neutral are clearly below the ones obtained on the data collected from typically developing children and seem to be the most difficult tasks.

CONCLUSIONS

Summing up, we first described the speech emotion database that has been used for evaluation; it is unique in composing speech data of children on the spectrum and a control group under the same conditions in three languages, namely English, Swedish, and Hebrew. Then, we discussed results concerning the classification of ASC children’s emotional expressions, evaluating each of the emotions against the remaining emotional states. Together with the classification evaluation, we analysed how classification performances change between the control and focus group. The results show common behaviour on the English and Swedish dataset, where the Angry task was found to be of poor performances close to chance level. On the other hand, in the Hebrew dataset, the most difficult task was Afraid. Looking at the deltas between the overall performances obtained within the two groups, we observed – on average – rather different results on the English and Hebrew datasets, whereas experiments on the Swedish dataset shows a smaller delta with similar performances. The caveat has to be made that this is a preliminary study, and deeper cross-language analysis is needed. Future experiments will address cross-corpus and feature analysis to investigate relevant language-independent acoustic features.

Acknowledgement

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 289021 (ASC-Inclusion).

ADDITIONAL AUTHORS

Sven Bölte, Steve Berggren, Daniel Lundqvist, Sigrid Elfström (Center of Neurodevelopmental Disorders, Karolinska Institute, Stockholm, Sweden, email: sven.bolte@ki.se).

REFERENCES

1. Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., and Narayanan, S. The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. *Journal of Speech, Language, and Hearing Research* 57, 4 (2014), 1162–1177.
2. Eyben, F., Wöllmer, M., and Schuller, B. openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor. In *Proceedings of the 18th ACM International Conference on Multimedia, MM 2010*, ACM, ACM (Florence, Italy, October 2010), 1459–1462.
3. Golan, O., and Baron-Cohen, S. Systemizing empathy: Teaching adults with asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Development and Psychopathology* 18, 02 (2006), 591–617.
4. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The weka data mining software: an update. *SIGKDD Explor. Newsl.* 11, 1 (Nov. 2009), 10–18.
5. Johnstone, T., van Reekum, C. M., Hird, K., Kirsner, K., and Scherer, K. R. Affective speech elicited with a computer game. *Emotion* 5, 4 (2005), 513.
6. Jones, C., and Sutherland, J. Acoustic emotion recognition for affective computer gaming. In *Affect and Emotion in Human-Computer Interaction*, C. Peter and R. Beale, Eds., vol. 4868 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2008, 209–219.
7. Kim, J., Bee, N., Wagner, J., and André, E. Emote to win: Affective interactions with a computer game agent. *GI Jahrestagung 1* (2004), 159–164.
8. Marchi, E., Batliner, A., Schuller, B., Fridenzon, S., Tal, S., and Golan, O. Speech, Emotion, Age, Language, Task, and Typicality: Trying to Disentangle Performance and Feature Relevance. In *Proceedings First International Workshop on Wide Spectrum Social Signal Processing (WS3P 2012)*, ASE/IEEE SocialCom 2012, ASE/IEEE, IEEE (Amsterdam, The Netherlands, September 2012).
9. Marchi, E., Ringeval, F., and Schuller, B. Perspectives and Limitations of Voice-controlled Assistive Robots for Young Individuals with Autism Spectrum Condition. In *Speech and Automata in Health Care (Speech Technology and Text Mining in Medicine and Healthcare)*, A. Neustein, Ed. De Gruyter, Berlin, 2014. invited contribution.
10. Marchi, E., Schuller, B., Batliner, A., Fridenzon, S., Tal, S., and Golan, O. Emotion in the Speech of Children with Autism Spectrum Conditions: Prosody and Everything Else. In *Proceedings 3rd Workshop on Child, Computer and Interaction (WOCCI 2012)*, *Satellite Event of INTERSPEECH 2012*, ISCA, ISCA (Portland, OR, September 2012).
11. Mower, E., Black, M. P., Flores, E., Williams, M., and Narayanan, S. Rachel: Design of an emotionally targeted interactive agent for children with autism. In *International Conference on Multimedia Computing and Systems/International Conference on Multimedia and Expo* (2011), 1–6.
12. Potamianos, A., and Narayanan, S. A review of the acoustic and linguistic properties of children’s speech. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on* (Oct 2007), 22–25.
13. Schuller, B., Marchi, E., Baron-Cohen, S., O’Reilly, H., Pigat, D., Robinson, P., Davies, I., Golan, O., Fridenzon, S., Tal, S., Newman, S., Meir, N., Shillo, R., Camurri, A., Piana, S., Stagliano, A., Bölte, S., Lundqvist, D., Berggren, S., Baranger, A., and Sullings, N. The state of play of ASC-Inclusion: An Integrated Internet-Based Environment for Social Inclusion of Children with Autism Spectrum Conditions. In *Proceedings 2nd International Workshop on Digital Games for Empowerment and Inclusion (IDGEI 2014)*, L. Paletta, B. Schuller, P. Robinson, and N. Sabouret, Eds., ACM, ACM (Haifa, Israel, February 2014). 8 pages, held in conjunction with the 19th International Conference on Intelligent User Interfaces (IUI 2014).
14. Schuller, B., Marchi, E., Baron-Cohen, S., O’Reilly, H., Robinson, P., Davies, I., Golan, O., Fridenzon, S., Tal, S., Newman, S., Meir, N., Shillo, R., Camurri, A., Piana, S., Bölte, S., Lundqvist, D., Berggren, S., Baranger, A., and Sullings, N. ASC-Inclusion: Interactive Emotion Games for Social Inclusion of Children with Autism Spectrum Conditions. In *Proceedings 1st International Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI 2013) held in conjunction with the 8th Foundations of Digital Games 2013 (FDG)*, B. Schuller, L. Paletta, and N. Sabouret, Eds., ACM, SASDG (Chania, Greece, May 2013).
15. Schuller, B., Steidl, S., Batliner, A., Nöth, E., Vinciarelli, A., Burkhardt, F., van Son, R., Weninger, F., Eyben, F., Bocklet, T., Mohammadi, G., and Weiss, B. The INTERSPEECH 2012 Speaker Trait Challenge. In *Proceedings INTERSPEECH 2012, 13th Annual Conference of the International Speech Communication Association*, ISCA, ISCA (Portland, OR, September 2012). 4 pages.
16. Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Weninger, F., Eyben, F., Marchi, E., Mortillaro, M., Salamin, H., Polychroniou, A., Valente, F., and Kim, S. The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism. In *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*, ISCA, ISCA (Lyon, France, August 2013). 5 pages.
17. Schuller, B., Vlasenko, B., Eyben, F., Wöllmer, M., Stuhlsatz, A., Wendemuth, A., and Rigoll, G. Cross-Corpus Acoustic Emotion Recognition: Variances and Strategies. *IEEE Transactions on Affective Computing* 1, 2 (July-December 2010), 119–131.
18. Yildirim, S., Narayanan, S., and Potamianos, A. Detecting emotional state of a child in a conversational computer game. *Computer Speech and Language* 25, 1 (2011), 29 – 44.