INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Automatic Analysis of Typical and Atypical Encoding of Spontaneous Emotion in the Voice of Children

Fabien Ringeval[1,3,4], *Erik Marchi*[1,3], *Charline Grossard*[5],
Jean Xavier[5], *Mohamed Chetouani*[5], *David Cohen*[5], *Björn Schuller*[1,2,3]

[1]audEERING GmbH, Gilching, Germany
[2]Department of Computing, Imperial College London, UK
[3]Chair of Complex & Intelligent Systems, University of Passau, Germany
[4]Laboratoire d'Informatique de Grenoble, Université Grenoble Alpes, France
[5]Institute of Intelligent Systems and Robotics, Université Pierre et Marie Curie, Paris, France

fabien.ringeval@uni-passau.de

Abstract

Children with Autism Spectrum Disorders (ASD) present significant difficulties to understand and express emotions. Systems have thus been proposed to provide objective measurements of acoustic features used by children suffering from ASD to encode emotion in speech. However, only a few studies have exploited such systems to compare different groups of children in their ability to express emotions, and even less have focused on the analysis of spontaneous emotion. In this contribution, we provide insights by extensive evaluations carried out on a new database of spontaneous speech inducing three emotion categories of valence (positive, neutral, and negative). We evaluate the potential of using an automatic recognition system to differentiate groups of children, i.e., pervasive developmental disorders, pervasive developmental disorders not-otherwise specified, specific language impairments, and typically developing, in their abilities to express spontaneous emotion in a common unconstrained task. Results show that all groups of children can be differentiated directly (diagnosis recognition) and indirectly (emotion recognition) by the proposed system.

Index Terms: affective computing, spontaneous emotions, autism spectrum disorders, language impairments

1. Introduction

The ability to communicate with speech requires the acquisition of codes that link acoustic realisation to both linguistic [1] and socio-affective related meanings [2, 3]. The acquisition and correct use of such codes, which are supposed to be functional in the early stages of a child's life [4], play an essential role in the inter-subjective development and social interaction abilities of children. As a consequence, most children presenting speech or developmental disorders have limited social interactions, which contributes to social isolation [5].

International classifications differentiate Specific Language Impairment (SLI) from those that are symptomatic of a developmental disorder, such as ASD. The former can affect both expressive and receptive language and is defined as a 'pure' language impairment. The latter, ASD, is characterised by severe deficits and pervasive impairment in several areas of development such as reciprocal social interactions, communication skills and stereotyped behaviours, interests and activities [6]. Because of the clinical heterogeneity of ASD, the recent DSM-5 decided to adopt a single diagnosis and to specify some dimensional features [7]. The DSM-4 distinguished ASD subtypes [8]: e.g. Autism Disorders (AD), with symptoms in all areas that characterise Pervasive Developmental Disorders (PDD); or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS), which is characterised by social, communicative and/or stereotypic impairments that are less severe than in AD and appear later in life [7]. Some authors have shown that AD, PDD-NOS and SLI had relied on different linguistic mechanisms and that expressive syntax, pragmatic skills, and some intonation features could be considered as language differential markers of pathology [9, 10].

Both clinicians and researchers are facing a huge increase in the prevalence of ASD, resulting from the expansion of the diagnostic criteria, but also from a better awareness of the condition and the acceptance that ASD is a lifelong condition [11]. Recently, systems have been proposed to provide objective measurements of acoustic features used by children suffering from ASD to encode non-verbal information in speech by the use of prosody [10, 12, 13, 14]. Analyses can be then performed indirectly, by assessing the performance of a child on a given task, e.g., producing specific prosodic contours to convey sentence modality [10], or emotion [12, 14]. In this case, the system is tuned for each group of children, e.g., typically developing (TD) and ASD, and performance can be compared between the groups to provide cues regarding the observed atypicalities of ASD. Analyses can also be performed directly, as an automatic diagnosis, by comparing the children's groups in the task [15, 16]. In this case, the system is tuned to search for differences in speech production between each group of children, which can also be a mean to identify the particularities of ASD.

Such systems can be used to help clinicians to improve the diagnosis, but also to develop tools based on information and communication technology, which enable users to access professional support on-line [18]. Recent studies have indeed shown that it is possible to improve emotional skills of ASD children in both emotion perception and production, by providing them interactive tools integrating affective computing [18]. However, the automatic processing of children's speech is challenging, as they present significant differences compared to the voice of adults [19], and even more when they are affected by ASD.

1.1. Contribution of this work

The present study focuses on the recognition of spontaneous emotional expressions in the voice of AD, PDD-NOS, SLI, and

http://dx.doi.org/10.21437/Interspeech.2016-766

Figure 1: Extracts from the book "Frog where are you?" [17] that were used for recording spontaneous emotional speech production from children. We hypothesised that images with *Neutral* valence (left picture, the beginning of the story: the boy is sleeping in his room while his frog is escaping), *Negative* valence (middle picture, the middle of the story: the boy has been captured by a deer while searching for his frog) and *Positive* valence (right picture, the end of the story: the boy finally finds his frog) correlated with emotion production.

Feature set	Spectral	Source	Duration	Total
eGeMAPS	48	48	6	102
IS09	336	48	0	384
IS10	1216	212	154	1582
IS11	2808	272	1288	4368
ComParE	4366	397	1610	6373

Table 1: Distribution of acoustic features in spectral/energy-related, source/excitation-related and duration-related feature sets for different configurations of openSMILE.

Feature set	Negative	Neutral	Positive	All
a. Typicality				
eGeMAPS	83.1	82.6	79.0	83.2
IS09	83.2	80.5	79.7	84.0
IS10	87.1	85.2	85.2	87.6
IS11	**89.1**	**87.5**	**87.4**	**89.4**
ComParE	88.0	85.1	86.5	86.3
b. Diagnosis				
eGeMAPS	46.2	48.2	44.9	48.2
IS09	50.8	45.8	46.7	51.2
IS10	51.8	47.1	50.1	53.7
IS11	**54.7**	**52.2**	**51.3**	**56.4**
ComParE	53.2	51.2	50.9	56.2

Table 2: UAR – typicality (2 classes) or diagnosis (4 classes) recognition for each feature set and emotion category.

typically developing (TD) children. We investigate the classification performances with expert-based reduced feature sets against large sets of features that include a vast number of spectral-, source-, and duration-related features. This study further focuses on the automatic discrimination of typicality between TD children and children suffering from AD, PDD-NOS, and SLI. For this purpose, a new database – Child Pathological and Emotional Speech database (CPESD) – is introduced, which will be made available to academic researchers. To the best knowledge of the authors, this is the very first study that performs automatic analysis of spontaneous emotion in children's speech with AD, PDD-NOS, and SLI. We hypothesised that we could distinguish: (i) different types of emotion based on prosodic features; (ii) patients with speech impairments from TD controls; (iii) different profiles within patients with speech impairments since some authors suggested that AD, SLI, and/or PDD-NOS may have different profile of emotional impairments [20, 21, 22].

2. Child Pathological & Emotional Speech Database

We received approval by the Ethical Committee of the Pitié-Salpétrière Hospital to conduct recruitment and speech recording of children. Consents were obtained from parents or legal caregivers of all participants. Thirty-five monolingual participants with communicative verbal skills were recruited in two university departments of child and adolescent psychiatry located in Paris, France. They consulted for ASD and/or SLI which were diagnosed as AD, PDD-NOS, or SLI, according to DSM IV criteria [8]. Patients were matched for age, sex, academic grades, and lexical abilities. Socio-demographic and clinical characteristics of the participants are available in [10]. We also recruited a group of 70 TD children matched for age and sex (1 patient for 2 TD) in elementary schools. A questionnaire was used to exclude children with learning disorders, an history of speech, language, hearing, or general learning problems.

Our main goal was to compare children's abilities to use prosody to encode pragmatic and affect in speech. A first task was based on the reproduction of intonation contour and was analysed in a previous study [10]. The second task was based on a story telling of a pictured book "Frog where are you?" [17], wherein a little boy tries to find his escaped frog during the night. The task was originally developed to assess language production in a standardised but unconstrained manner. Here, we assume that the child is supposed to produce prosodic cues during the story telling that are correlated to the levels of the emotional valence, which was categorised in three categories by a psychologist: Negative/Neutral/Positive. In total, the pictured book included 15 emotionally negative, 6 emotionally neutral and 5 emotionally positive pictures, cf. Figure 1. Three pictures considered ambivalent because of ambiguous interpreta-

Group	Valence										
	Negative			Neutral			Positive			All	
	#	%	Duration	#	%	Duration	#	%	Duration	#	Duration
AD	335	35.4^{*N}	$2.13^{*N,T}$	137	34.8	2.40^{*T}	94	29.8^{*N}	$2.43^{*N,S,T}$	566	2.25^{*T}
NOS	283	$30.1^{*A,T}$	$2.31^{*A,S,T}$	126	32.2^{*T}	2.51^{*T}	118	$37.7^{*A,S,T}$	$2.08^{*A,T}$	527	$2.31^{*S,T}$
SLI	530	31.8	$2.17^{*N,T}$	243	35.0	2.30^{*T}	184	$33.2^{*N,T}$	$2.16^{*A,T}$	957	$2.20^{*N,T}$
TD	2146	33.8^{*N}	$2.83^{*A,N,S}$	970	36.7^{*N}	$2.89^{*A,N,S}$	623	$29.5^{*N,S}$	$2.78^{*A,N,S}$	3739	$2.84^{*A,N,S}$

Table 3: Number, relative proportion, and mean duration of utterances per emotion class and group. $* = p < 0.05$: alternative hypothesis is true when comparing data between children groups, i. e., A, N, S, and T; AD (A): autism disorders; NOS (N): pervasive developmental disorders not-otherwise specified; SLI (S): specific language impairment; TD (T): typically developing.

	eGeMAPS			IS09		IS10			IS11			ComParE		
Task	Spec.	Sour.	Dur.	Spec.	Sour.	Spec.	Sour.	Dur.	Spec.	Sour.	Dur.	Spec.	Sour.	Dur.
Typ.	83.4	71.1	66.7	82.0	74.8	86.8	76.3	74.9	**88.3**	**80.6**	**85.7**	88.2	72.8	84.4
Diag.	46.3	39.4	34.1	50.0	40.1	53.3	42.0	41.3	**56.7**	**43.5**	**49.5**	56.5	38.9	47.9

Table 4: UAR – typicality (2 classes) or diagnosis (4 classes) recognition for each feature subset and all emotion categories.

tion were excluded.

We collected nearly 10 hours of recording: 7 h 38 min for TD children, 1 h 35 min for children with AD, 1 h 12 min for children with PDD-NOS, and 1 h 56 min for children with SLI. Recordings were then segmented automatically into groups of breaths, using the energy contour. As many sources of perturbation appeared during the recordings (e. g., false-starts, repetitions or noise from the environment), the obtained speech segments were further manually processed; only utterances that had a complete prosodic contour, i. e., whatever the pronounced words, were kept. Statistics (number, relative proportion, and mean duration) on those utterances are provided for each emotion category, and all, in Table 3. Those data already provide some interesting insights: all TD children produced utterances which are significantly longer than AD, PDD-NOS, and SLI children for all emotion categories ($p < 0.5$, two-tailed t test); we observed the opposite on the constrained task of intonation contour imitation [10]. Moreover, spontaneous speech production of PDD-NOS children focused significantly more on positive emotions compared to all other groups ($p < 0.5$).

3. Experiments

Two main tasks were performed: automatic recognition of typicality (direct analysis), and emotion (indirect analysis). The typicality task concerns the classification of TD children vs all other children. Additionally, we performed the classification of each group of children (diagnosis). The emotion task covers the recognition of the three classes of emotional valence, i. e., positive, neutral, and negative. This task was performed either on each group separately, or with models trained on TD children.

3.1. Acoustic features

Acoustic features were automatically extracted from the speech waveform on the utterances by using our open-source openSMILE feature extractor in its recent 2.2 release [23]. Five different feature sets were investigated: large brute-forced feature sets (IS09, IS10, IS11, and ComParE), which have all been used for paralinguistic information retrieval, and a smaller, expert knowledge based feature set (eGeMAPS). Those feature sets cover spectral-, source- and duration-related feature space with

Feature set	AD	NOS	SLI	TD
eGeMAPS	36.2	35.6	**42.7**	44.1
IS09	35.5	39.3	39.3	42.9
IS10	37.7	**40.5**	42.0	43.1
IS11	**38.0**	37.9	40.9	**44.5**
ComParE	35.8	37.0	38.9	42.3

Table 5: UAR – spontaneous emotional valence recognition (3 classes) for each feature set and groups of children; AD: autism disorders; NOS: pervasive developmental disorders not-otherwise specified; SLI: specific language impairment; TD: typically developing.

different levels of detail, cf. Table 1. The first four sets, i. e., IS09, IS10, IS11, and ComParE, show a clear tendency in enlarging the feature space over the years, by including further low-level acoustic descriptors and associated functionals. Recently, this "brute-forcing" approach has been revisited, with investigations on a small, expert knowledge based feature set, eGeMAPS [24]. A detailed description and implementation of these feature sets, which is impossible to provide here, is given in [25].

3.2. Setup and evaluation

We used Support Vector Machines (SVMs) for the classification tasks with LIBSVM [26], as they are a well known standard method that can handle both high and low dimensional data. The SVM training has been made with the three different kernels: linear, polynomial (3rd order), and Gaussian (γ parameter was set to default value); the complexity parameter was set to default value ($C = 1$). Results are always presented with the best kernel. To ensure speaker independent evaluations, we performed a Leave-One-Speaker-Out (LOSO) cross-validation in all experiments. Because all data sets are unbalanced, we applied upsampling of the under-represented classes in all the evaluation experiments. For the same reason, we used the unweighted average recall (UAR) of the classes as scoring metric. Standardisation of the features, i. e., feature val-

	eGeMAPS			IS09		IS10			IS11			ComParE		
Group	Spec.	Sour.	Dur.	Spec.	Sour.	Spec.	Sour.	Dur.	Spec.	Sour.	Dur.	Spec.	Sour.	Dur.
AD	34.8	**37.4**	**37.7**	37.0	37.4	35.9	34.8	33.6	**39.8**	37.3	36.8	38.7	33.5	35.8
NOS	38.5	**39.09**	30.2	38.4	37.1	**41.4**	36.5	**39.15**	34.1	36.9	37.0	33.4	34.98	37.0
SLI	**43.7**	**38.5**	38.0	38.5	37.1	42.1	38.2	34.9	40.40	37.4	38.7	40.5	37.5	**38.9**
TD	**46.0**	**41.3**	34.6	43.1	36.0	43.7	38.3	38.9	44.4	38.3	42.0	44.2	39.9	**42.3**

Table 6: UAR – spontaneous emotion recognition (3 classes) for each feature subset and groups of children; AD: autism disorders; NOS: pervasive developmental disorders not-otherwise specified; SLI: specific language impairment; TD: typically developing.

Feature set	AD	NOS	SLI	TD
Spectral + Source	40.0	39.7	43.3	44.7
Source + Duration	37.3	37.3	39.5	42.8
Spectral + Duration	**40.6**	41.5	39.0	43.2
Spec. + Sour. + Dur.	39.8	**42.1**	39.1	43.2
Best group	39.8	41.4	**43.7**	**46.0**

Table 7: UAR – spontaneous emotion recognition with 3 classes, combination of the best feature subsets; AD: autism disorders; NOS: pervasive developmental disorders not-otherwise specified; SLI: specific language impairment; TD: typically developing.

ues are normalised to zero-mean and unit standard deviation, was performed for each speakers for all emotion recognition tasks. Whereas for both typicality and diagnosis tasks, we standardised the features of all speakers with the on-line approach, i. e., mean and standard-deviation were computed on the training partition and applied on both training and testing data.

4. Results

4.1. Typicality & Diagnosis

Results obtained on typicality and diagnosis are given for each emotion class and each feature set in Table 2. One may note that all obtained performance are far above the chance level, in agreement with [14], despite being lower than on the constrained task, i. e., intonation contour imitation [27]. In order to gain further insights, we performed automatic recognition of typicality and diagnosis with each different feature subset, i. e., spectral-, source-, and duration-related features. Results show that, spectral-related features are the most contributing in the two tasks, and can perform even better when taken alone for diagnosis, cf. Table 4.

4.2. Emotion

Results obtained on the automatic recognition of the emotional valence are given in Table 5. Our hypothesis that the spontaneous description of the pictured book will be correlated with the emotional valence depicted in the images is validated by the experiments, because the system performs significantly better than chance for all groups of children. Obviously, TD children obtained the best performance, and all other children obtained a significantly lower performance ($p < 0.5$); two-tailed t test. A detailed analysis of each feature subset shows that, (i) spectral-related features provide the best performance for all groups of children, and (ii) the minimalistic feature set, i. e., eGeMAPS, performed remarkably well on all groups, especially for source-related features, cf. Table 6. Results obtained with different combinations of the best feature subsets show that, the performance was improved further for the PDDs, by combining spectral- and duration-related features for AD, and all feature subsets for PDD-NOS, cf. Table 7.

Feature set	AD	NOS	SLI	TD
Spectral	35.8	**41.7**	35.8	46.0
Source	**39.3**	38.4	33.9	41.3
Duration	34.2	37.8	**38.0**	42.3
Spectral + Source	38.0	**43.5**	33.2	44.7
Source + Duration	36.3	**39.2**	37.5	42.8
Spectral + Duration	34.7	**40.5**	38.7	43.2
Spec. + Sour. + Dur.	35.1	**40.7**	38.2	43.2
IS11	38.3	**40.8**	39.6	44.5

Table 8: UAR – spontaneous emotional valence recognition with 3 classes and model training on TD; AD: autism disorders; NOS: pervasive developmental disorders not-otherwise specified; SLI: specific language impairment; TD: typically developing.

Finally, in order to investigate how the models obtained on TD children could generalise on the others groups of children, we performed a mismatched evaluation, by training models on TD and testing on AD, PDD-NOS, and SLI. Results show that, there are some specific associations between feature space and pathology; models obtained from TD children generalised best on AD with source-related features, PDD-NOS with spectral-related features, and SLI with duration-related features. Moreover, PDD-NOS children obtained systematically the best performance for all combinations of the feature subsets, cf. Table 8. This result supports the hypothesis that PDD-NOS may express more emotion than AD [20, 21, 22].

5. Conclusions

A new speech database of spontaneous emotions produced by AD, PDD-NOS, SLI and TD speaking children has been introduced: CPESD. Extensive experiments have been performed on this database, showing that all groups of children can be differentiated directly (typicality, diagnosis) and indirectly (emotion) with an automatic recognition system. A detailed analysis has shown that, large scale features are necessary to differentiate TD from the other groups, whereas the eGeMAPS set provides the best results for the emotion recognition task.

6. Acknowledgements

The research leading to these results has been funded by the European Community's Seventh Framework Programme through the ERC Starting Grant No. 338164 (iHEARu), and the European's Union's Horizon 2020 Programme through the Research Innovative Action #688835 (DE-ENIGMA).

7. References

[1] D. Crystal, "The linguistic status of prosodic and paralinguistic features," *Proc. of the University of Newcastle-upon Tyne Philosophical Society*, vol. 1, pp. 93–108, 1966.

[2] ——, "Prosodic and paralinguistic correlates of social categories," in *Social anthropology*, E. Ardener, Ed. London: Tavistock, 1971, pp. 185–206.

[3] E. Marchi, Y. Zhang, F. Eyben, F. Ringeval, and B. Schuller, "Autism and Speech, Language, and Emotion – a Survey," in *Evaluating the Role of Speech Technology in Medical Case Management*, H. Patil and M. Kulshreshtha, Eds. Berlin: De Gruyter, 2015, invited contribution, to appear.

[4] P. K. Kuhl, "Early language acquisition: cracking the speech code," *Nature Reviews Neuroscience*, vol. 5, pp. 831–843, November 2004.

[5] E. Müller, A. Schuler, and G. B. Yates, "Social challenges and supports from the perspective of individuals with Asperger syndrome and other autism spectrum disabilities," *Autism*, vol. 12, no. 2, pp. 173–190, March 2008.

[6] I. Rapin and D. A. Allen, "Developmental language: nosological consideration," in *Neuropsychology of Language, Reading and Spelling*, V. Kvik, Ed. Washington, DC: New York: Academic Press, 1983.

[7] J. Xavier, C. Bursztejn, M. Stiskin, R. Canitano, and D. Cohen, "Autism spectrum disorders: An historical synthesis and a multidimensional assessment toward a tailored therapeutic program," *Research in Autism Spectrum Disorders*, vol. 18, pp. 21–23, October 2015.

[8] "Diagnostic and Statistical Manual of mental disorders (4th Ed.)." Washington, DC: American Psychiatric Association, 1994.

[9] J. Demouy, M. Plaza, J. Xavier, F. Ringeval, M. Chetouani, D. Périsse, D. Chauvin, S. Viaux, B. Gols, D. Cohen, and L. Robel, "Differential language markers of pathology in Autism, Pervasive Developmental Disorder Not Otherwise Specified and Specific Language Impairment," *Research in Autism Spectrum Disorders*, vol. 5, no. 4, pp. 1402–1412, October-December 2011.

[10] F. Ringeval, J. Demouy, G. Szaszák, M. Chetouani, L. Robel, J. Xavier, D. Cohen, and M. Plaza, "Automatic intonation recognition for prosodic assessment of language impaired children," *IEEE Transactions on Audio, Speech & Language Processing*, vol. 19, no. 5, pp. 1328–1342, July 2011.

[11] J. L. Matson and A. M. Kozlowski, "The increasing prevalence of autism spectrum disorders," *Research in Autism Spectrum Conditions*, vol. 5, no. 1, pp. 418–425, Januray–March 2011.

[12] E. Marchi, B. Schuller, A. Batliner, S. Fridenzon, S. Tal, and O. Golan, "Emotion in the Speech of Children with Autism Spectrum Conditions: Prosody and Everything Else," in *Proc. of WOCCI 2012*, ISCA. Portland, OR: ISCA, September 2012.

[13] D. Bone, C.-C. Lee, M. P. Black, M. E. Williams, S. Lee, P. Levitt, and S. Narayanan, "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 4, pp. 1162–1177, August 2014.

[14] E. Marchi, B. Schuller, S. Baron-Cohen, O. Golan, S. Bölte, P. Arora, and R. Häb-Umbach, "Typicality and Emotion in the Voice of Children with Autism Spectrum Condition: Evidence Across Three Languages," in *Proc. of INTERSPEECH 2015*, ISCA. Dresden, Germany: ISCA, September 2015, pp. 115–119.

[15] E. Marchi, F. Ringeval, and B. Schuller, "Voice-enabled assistive robots for handling autism spectrum conditions: an examination of the role of prosody," in *Speech and Automata in Health Care (Speech Technology and Text Mining in Medicine and Healthcare)*, A. Neustein, Ed. Boston/Berlin/Munich: De Gruyter, 2014, pp. 207–236.

[16] E. Marchi, A. Batliner, B. Schuller, S. Fridenzon, S. Tal, and O. Golan, "Speech, Emotion, Age, Language, Task, and Typicality: Trying to Disentangle Performance and Feature Relevance," in *Proc. WS^{3}P 2012, SocialCom 2012*, ASE/IEEE. Amsterdam, The Netherlands: IEEE, September 2012, 8 pages.

[17] M. Mayer, *Frog where are you?* New York: Dial Books for young readers, 1969.

[18] B. Schuller *et al.*, "Recent developments and results of ASC-Inclusion: An Integrated Internet-Based Environment for Social Inclusion of Children with Autism Spectrum Conditions," in *Proc. of IDGEI 2015*. Atlanta, GA: ACM, March 2015.

[19] A. Potamianos and S. Narayanan, "A review of the acoustic and linguistic properties of childrens speech," in *Proc. of the IEEE 9th Workshop on Multimedia Signal Processing*, October 2007, pp. 22–25.

[20] C. A. Mazefsky, J. Herrington, M. Siegel, A. Scarpa, B. B. Maddox, L. Scahill, and S. W. White, "The Role of Emotion Regulation in Autism Spectrum Disorder," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 52, no. 7, pp. 679–688, July 2013.

[21] J. Xavier, L. Vannetzel, S. Viaux, A. Leroy, M. Plaza, S. Tordjman, C. Mille, C. Bursztejn, D. Cohen, and J.-M. Guile, "Reliability and diagnostic efficiency of the Diagnostic Inventory for Disharmony (DID) in Youths with Pervasive Developmental Disorder and Multiple Complex Developmental Disorder," *Research in Autism Spectrum Disorders*, vol. 5, no. 4, pp. 1493–1499, October-December 2011.

[22] M.-C. Lai, M. V. Lombardo, B. Chakrabarti, and S. Baron-Cohen, "Subgrouping the Autism Spectrum: Reflections on DSM-5," *PLoS Biology*, vol. 11, no. 4, pp. 21–23, April 2013, e1001544.

[23] F. Eyben, F. Weninger, F. Groß, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in *Proc. of the 21st ACM International Conference on Multimedia (ACM MM)*, Barcelona, Spain, October 2013, pp. 835–838.

[24] F. Eyben, K. Scherer, B. Schuller, J. Sundberg, E. André, C. Busso, L. Devillers, J. Epps, P. Laukka, S. Narayanan, and K. Truong, "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," *IEEE Transactions on Affective Computing*, 2015, 14 pages, to appear.

[25] F. Eyben, "Real-time speech and music classification by large audio feature space extraction," Ph.D. dissertation, Technische Universität München, 2014, springer.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, April 2010, Article No. 27.

[27] B. Schuller *et al.*, "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in *Proc. of INTERSPEECH 2013*, Lyon, France, August 2013, pp. 148–152.