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

Does She Speak RTT? Towards an Earlier Identification of Rett Syndrome Through Intelligent Pre-linguistic Vocalisation Analysis

Florian B. Pokorny[1,2,3], *Peter B. Marschik*[1,3,4], *Christa Einspieler*[1], *Björn W. Schuller*[5,6]

[1]Research Unit iDN – interdisciplinary Developmental Neuroscience, Institute of Physiology, Center for Physiological Medicine, Medical University of Graz, Austria
[2]Machine Intelligence & Signal Processing group, Technische Universität München, Germany
[3]Brain, Ears & Eyes – Pattern Recognition Initiative (BEE-PRI), BioTechMed-Graz, Austria
[4]Centre of Neurodevelopmental Disorders (KIND), Karolinska Institutet, Stockholm, Sweden
[5]Chair of Complex & Intelligent Systems, University of Passau, Germany
[6]Machine Learning Group, Department of Computing, Imperial College London, UK
florian.pokorny@medunigraz.at

Abstract

For many years, an apparently normal early development has been regarded as a main characteristic of Rett syndrome (RTT), a severe progressive neurodevelopmental disorder almost exclusively affecting girls/females. The speech-language domain represents a key domain for the clinical diagnosis of RTT, which usually happens around three years of age. Recent studies have built upon the assumption that this domain is already affected in the prodromal period. Aiming to find RTT-specific speech-language atypicalities on signal level as early acoustic markers, we analysed more than 16 hours of home video recordings of 4 girls later diagnosed with RTT and 4 typically developing girls aged 6 to 12 months. We segmented a total of 4 678 pre-linguistic vocalisations. A comprehensive set of acoustic features was extracted from the vocalisations as basis for the classification paradigm RTT versus typical development. A promising mean unweighted recognition accuracy of 76.5% was achieved using linear kernel support vector machines and 4-fold leave-one-speaker-pair-out cross-validation. To the best of our knowledge, this is the first approach to automatically identify infants later diagnosed with RTT based on acoustic characteristics of pre-linguistic vocalisations. Our findings may build the basis for facilitating earlier identification and thus an avenue for an earlier entry into intervention.

Index Terms: Rett syndrome, early detection, infant vocalisation analysis, speech-language pathology

1. Introduction

Rett syndrome (RTT) is a severe neurodevelopmental disorder almost exclusively occurring in females [1, 2] with a prevalence of 1:5 000 to 1:10 000 live female births (rare disease) [3]. It was first described in 1966 by the Austrian neuropaediatrician Andreas Rett [4]. More than 30 years later, mutations in the X-linked gene encoding Methyl-CpG-binding protein 2 (MeCP2) were identified as the cause for RTT in most (but not necessarily in all) cases [5]. RTT has long been believed to be characterised by an apparently normal early development followed by a period of profound neurological regression. This regression affects – among other neurodevelopmental functions such as purposeful hand use or cognitive skills – the use of expressive language [6, 7, 8, 2, 9]. Thus, the detection of the speech language dysfunctions is one of the key domains for the diagnosis of RTT, which still is a clinical diagnosis at first, confirmed by genetic testing [2]. Recent observational studies have found increasing evidence that this domain is already affected in the pre-regression period, before diagnosis, challenging the paradigm of normal early development (e. g., [10, 11, 12, 13, 14]).

Besides delays in the achievement of certain speech-language milestones, or even their non-achievement in girls with RTT (e. g., [13]), verbal characteristics of infant vocalisations have been found to bear qualitative atypicalities (e. g., [11]). For six females with RTT, Marschik et al. [11] reported three different atypical vocalisation characteristics, of (i) pressed, (ii) inspiratory, and (iii) high-pitched crying-like quality. Interestingly, these vocalisations appeared intermittently with typical vocalisations as early as from the 7th month of life onwards. Marschik et al. further stated: "The intermittent character of normal versus abnormal behaviors might contribute to an early identification of children with possible genetic mutations, and provides evidence that speech-language functions are abnormal from the very beginning." [11, p.1]

In this study, we took on the challenge to start a very first attempt to itemise potential RTT-specific speech-language atypicalities on signal level. We aimed to make a step towards enabling an earlier identification of RTT on the basis of objective acoustic signal parameters in pre-linguistic vocalisations using machine learning methodology in order to facilitate an earlier entry into intervention for individuals with RTT.

2. Can we hear RTT?

In a case study (pilot listening experiment [Pokorny et al. 2016, in preparation]), we aimed to quantify the intermittent character of typical and atypical early vocalisations in RTT by means of listeners' vocalisation assessments. Therefore, we presented more than 300 pre-linguistic vocalisations of a girl later diagnosed with RTT separately to five professionals in the fields of speech-language pathology, developmental psychology, and/or developmental physiology. We asked the experts to rate whether they perceived the vocalisations typical or atypical. Atypical vocalisations should be qualitatively further classified into: rhythm, timbre, pitch, or other acoustic parameters as predominantly deviant feature (multiple answers were allowed).

About half of the vocalisations were rated atypical by at least one listener. Only nine vocalisations were consenta-

http://dx.doi.org/10.21437/Interspeech.2016-520

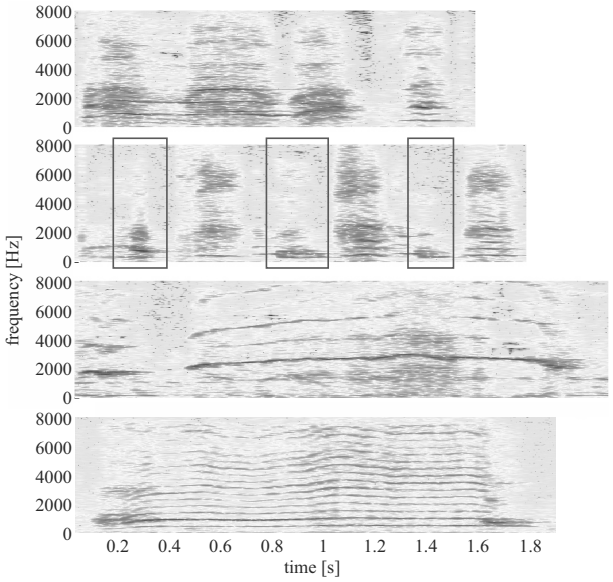

Figure 1: *Spectrograms of (i, top) a pressed babbling-like vocalisation, (ii, top middle) a vocalisation with distinctive inspiratory phases (marked with rectangular boxes), (iii, bottom middle) a high-pitched crying-like vocalisation, and (iv, bottom) a vocalisation without any atypical characteristics produced by a girl in the second half year of life later diagnosed with RTT.*

neously rated as atypical exhibiting characteristics as described in [11]. Figure 1 shows spectrograms of (i) a pressed vocalisation with inharmonic overtone structure rated as atypical in timbre by all five listeners, (ii) a vocalisation with distinctive inspiratory phases rated as atypical in timbre by four listeners and in rhythm by three listeners, (iii) a high-pitched crying-like vocalisation rated as atypical in pitch by four listeners, and (iv) a vocalisation with harmonic overtone structure rated as typical by all five listeners.

However, the listening experiment's low overall inter-rater reliability ($\kappa = 0.2$) indicated that atypicalities in early vocalisations may be hardly reliably identified by human listeners. This calls for machine-driven approaches to objectively define acoustic phenomena in early vocalisations in RTT.

3. Methods

In this study, we retrospectively focussed on pre-linguistic vocalisations of the second half year of life, as this striking and diagnostically relevant [15] period covers the transition from the use of first syllabic sounds and canonical babbling to the production of first meaningful words in typical development [16, 17, 18, 19, 20, 21].

3.1. Material

We reviewed more than 1 000 minutes of home video recordings of the second half year of life of 4 female infants later diagnosed with RTT and 4 typically developing (TD) female infants. The recordings were made by the infants' parents in typical family settings (e. g., playing situations, feeding, bathing) and during special family events (e. g., birthday parties). At the time of recording, the parents of the individuals later diagnosed with RTT were not aware of their daughters' medical condition. All eight participants stem from German-speaking families, who provided the audio-video material for the purpose of scientific analysis. The Institutional Review Board of the Medical University of Graz approved the method of retrospective audio-video analysis.

Table 1: *Age-specific distribution of the number of vocalisations segmented on the basis of footage of four infants later diagnosed with RTT and of four TD infants. ('-' indicates that no audio-video material was available.)*

Case	age [months]						Σ
#	7	8	9	10	11	12	
RTT1	-	363	273	323	180	243	1382
RTT2	-	-	-	150	-	-	150
RTT3	-	-	138	120	46	-	304
RTT4	73	26	57	73	68	66	363
Σ	73	389	468	666	294	309	2199
TD1	231	97	89	80	78	106	681
TD2	78	183	4	26	84	86	461
TD3	35	-	-	137	109	714	995
TD4	9	48	84	102	51	48	342
Σ	353	328	177	345	322	954	2479

3.2. Segmentation

Manual segmentation of vocalisations was carried out using the video coding tool Noldus Observer XT. A vocalisation was defined as an utterance underlying a vocal breathing group [22]. We did not include vocalisations that could not be ascribed to a video's participating infant with absolute certainty (e. g., in settings with more than one infant of about the same age present). We further excluded vegetative sounds such as breathing sounds, sneezes, hiccups, smacking sounds, etc. Relevant vocalisation detection in the video as well as raw segmentation were done by two female and two male research assistants following an intensive instruction by the first author. Prior to inclusion in this study, the first author verified each pre-selected vocalisation for validity and carried out the fine segmentation process. As itemised in Table 1, a total of 2 199 pre-linguistic vocalisations could be segmented within the footage of infants later diagnosed with RTT. The material of TD infants contained 2 479 pre-linguistic vocalisations. All segmented vocalisations were exported as audio tracks (44.1 kHz, 16 bit, 1 channel, PCM) for further analysis.

3.3. Analysis

To build the basis for vocalisation analysis/classification on signal level, feature extraction was carried out using the open-source tool kit openSMILE [23] in its current release [24]. Describing one of the most comprehensive standardised sets available, we extracted the official baseline features of the Interspeech Computational Paralinguistics Challenges (ComParE) used since 2013 (e. g., [25, 26]) from each vocalisation. The set comprises 6 373 higher-level features representing statistical functionals of a wide range of acoustic time-, spectral-, and/or energy-based short-term low-level descriptors' trajectories and their derivatives.

In order to investigate the binary classification paradigm RTT versus typical development, we split our data into training, development, and test partitions. Due to the small number of infants per class (4 RTT versus 4 TD), we decided for evaluating classification performance via a 4-fold leave-one-speaker-pair-

Table 2: *Allocation of four infants later diagnosed with RTT (RTT1–RTT4) and four TD infants (TD1–TD4) to training, development, and test partitions as well as class proportion of respective numbers of vocalisations/instances (RTT/TD) for 4-fold leave-one-speaker-pair-out cross-validation so as to create test partitions with minimal class imbalances. ('→' indicates that upsampling was performed due to a class imbalance in a training subset exceeding the ratio 3:2.)*

Run	Training	Develop	Test
1	RTT1,RTT2,TD2,TD3 1532/1456	RTT3,TD4 304/342	RTT4,TD1 363/681
2	RTT1,RTT4,TD3,TD4 1745/1337	RTT2,TD1 150/681	RTT3,TD2 304/461
3	RTT2,RTT3,TD1,TD4 454/1023 → 908/1023	RTT4,TD2 363/461	RTT1,TD3 1382/995
4	RTT3,RTT4,TD1,TD2 667/1142 → 1334/1142	RTT1,TD3 1382/995	RTT2,TD4 150/342

out cross-validation scheme. For each of the four validation runs the training subsets contained the vocalisations of two infants per class, the development and test subsets each contained the vocalisations of one infant per class. The vocalisations of each infant were included in the test partition exactly one time. To ensure maximal class balancing within the test subsets for each of the four validation runs, we pairwisely matched the numbers of segmented vocalisations per infant. Accordingly, the infant later diagnosed with RTT with the highest number of segmented vocalisations was grouped with the TD infant with the highest number of segmented vocalisations, etc. The detailed partitioning is given in Table 2. Training subsets with class imbalances exceeding the ratio 3:2 (training subset for third and fourth run, and training+development subset for third run) were upsampled applying (Weka's implementation of) the synthetic minority oversampling technique (SMOTE) [27].

To study the influence of infant-specific feature value distributions on classification performance, on the one hand, we (Method A) did not normalise/standardise feature values speaker/infant-dependently prior to classification. On the other hand, we (Method B) performed speaker/infant normalisation, i. e., all features were speaker/infant-dependently normalised to the interval [0,1], and finally, we (Method C) applied speaker/infant standardisation, i.e, all features were speaker/infant-dependently standardised to have zero mean and unit variance, before passing feature data to the classifier.

Known not to be sensitive to feature overfitting, we applied linear kernel support vector machines (SVMs) as classifier by means of the widely used data mining tool kit Weka [28]. The kernel complexity parameter C was optimised for each of the four validation runs within $\{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ on the basis of the respective development subset. Subsequently, the training subset and the development subset were combined to a final training subset for validation on the basis of the respective test subset for each of the four runs. For SVM training, we selected the sequential minimal optimisation algorithm.

4. Results

For Method A, i. e., classification without preceding speaker/infant normalisation/standardisation, we achieved a mean class unweighted recognition accuracy (UA) ± standard deviation of $59.4\% \pm 11.4\%$ over the four validation runs.

Table 3: *Classification results of 4-fold leave-one-speaker-pair-out cross-validation in form of class-specific numbers of test vocalisations (in-)correctly classified as class RTT or TD (confusion matrix), and mean and standard deviation (SD) of weighted and unweighted accuracies (WA and UA) for the case without speaker normalisation/standardisation (Method A) and the case with speaker normalisation (Method B). WA and UA are given in [%]. Values are rounded to one decimal place.*

	Method A		Method B	
$classified\ as \rightarrow$	RTT	TD	RTT	TD
RTT	1650	549	1586	613
TD	1239	1240	615	1864
	WA	UA	WA	UA
mean	62.6	59.4	79.1	76.5
SD	11.7	11.4	18.9	23.4

75.0% of the vocalisations of infants later diagnosed with RTT were correctly identified as class RTT, but the vocalisations of TD infants were classified about one half each as class TD and RTT. Method B, i. e., classification with preceding speaker/infant normalisation, performed significantly better (at a significance level of $\alpha = 0.001$ using a one-sided z-test) and reached a mean UA of $76.5\% \pm 23.4\%$. 72.1% of the vocalisations of infants later diagnosed with RTT and 75.2% of the vocalisations of TD infants were correctly classified. In two of the four validation runs weighted and unweighted accuracies higher than 90% were achieved. Method C, i. e., classification with preceding speaker/infant standardisation, performed with significantly lower accuracy (at a significance level of $\alpha = 0.001$ using a one-sided z-test) than the other two methods. Here, 63.3% of the vocalisations of infants later diagnosed with RTT were correctly identified, but also 64.9% of the vocalisations of TD infants were (incorrectly) assigned to the RTT class. None of the four validation runs using Method C exceeded the level of random guessing leading to a mean UA of $49.8\% \pm 0.9\%$. Detailed results for the effective methods A and B including both mean class weighted and unweighted accuracies and standard deviations, as well as overall confusion matrices are given in Table 3.

5. Discussion

On our dataset, the basic feasibility of an automatic recognition of pre-linguistic vocalisations produced by infants later diagnosed with RTT was supported. However, data preprocessing in terms of infant-specific data normalisation/standardisation prior to classification was a crucial step. On our data, classification with preceding infant normalisation significantly outperformed classification with preceding infant standardisation and classification neither with normalisation nor with standardisation. Cause for the significant differences among the three methods may be the intermittently occurring atypical vocalisations in infants later diagnosed with RTT that are, to date, hardly objectified/documented on signal level. Another influential role may play the inhomogeneity of our material with respect to audio quality and background noise in the home videos.

From a neurodevelopmental and linguistic point of view, we implemented a straight-forward top-down approach by not differentiating between (i) vocalisations produced in different developmental stages (we integrated all pre-linguistic vocalisations produced during the second half year of life to one dataset,

even though the exact age in months was known for each video clip), and (ii) different vocalisation types (e. g., quasi-resonant nuclei versus canonical babbling [16]) to a certain extent correlating with the developmental stages. Furthermore, we did not consider potential differences in the vocalisations' recording quality or possible stationary and/or transient background noise present in a vocalisation recording. However, with this study we intended to show basic feasibility and to lay the foundation of future approaches more and more modelling neurodevelopmental and linguistic knowledge to potentially increase recognition accuracy. To this end, a substantially higher amount of training data, thus, lots of more home video recordings of both infants later diagnosed with RTT and TD infants, will be necessary.

In general, the use of home video recordings in basic research involves certain risks and limitations. First of all, the videos were not recorded for the purpose of later scientific analysis. Therefore, situations are to a high degree not standardised. Furthermore, we can assume, that particular behaviours (e. g., the production of specific vocalisation types) are absent in an available dataset although they may be present in real life [29]. For example, an infant could hypothetically have already reached the speech-language milestone of canonical babbling, but within the scenes of the available video clips canonical babbling was never produced. Consequently, home videos do not allow for assessing frequencies of specific behaviours. Another aspect should be kept in mind, in particular when studying neurodevelopmental disorders based on home video material. Parents usually tend to stop recording when their infants start behaving obviously atypically or alarming, e. g., when producing peculiar vocalisation patterns. Finally, the presence of everyday background noise highly impedes acoustic analyses based on home video recordings, especially in small datasets. However, the low prevalence of RTT in combination with its current late mean age of diagnosis hampers the conductance of comprehensive prospective studies. Thus, at the moment, retrospective audio-video analysis is one of the best available approaches to 'look' or 'listen' back for studying early maldevelopment in infants with neurodevelopmental disorders normally not detected until toddlerhood, such as RTT [30]. Also in autism research, the retrospective analysis of home videos still represents a well-established approach [31, 32, 33].

Apart from the above discussed issues involved by the method of retrospective audio-video analysis, the impact of our study is limited due to the critically small number of infants per class, the variable number of segmented vocalisations per infant, the imbalanced distribution of available vocalisations per infant per month of life, and thereby, the imbalanced distribution of vocalisation types per infant available in the dataset. From the opposite point of view, it is worth mentioning, that promising recognition results could be achieved for training the classification model based on vocalisations of only three infants per class in each run. Anyhow, we are continuously expanding our database, but especially the acquisition of audio-video data of infants with rare neurodevelopmental disorders in the prodromal period is challenging and time-consuming. The data used for this study are rare and have been collected over years.

With regard to a possible future application of an approach like ours presented in this study for the purpose of assistance or decision support in clinical or paediatric settings, the intermittent character of typical and atypical vocalisations predominating the early speech-language development in individuals with RTT will prevent the possibility of reliably identifying an infant with RTT on the basis of just a few vocalisations. Besides a classification model trained on the basis of a high number of infants per class, for a reliable decision also a considerable number of test vocalisations of an infant with unknown outcome should be available to ensure that the data also contain a sufficient number of atypical vocalisations in case of RTT. In practical use, a prediction tool could be implemented working fully automatically on the basis of home video material, or, e. g., on the basis of audio material recorded 24 hours with a microphone attached to an infant's clothing. For such a tool, at least a reliable infant voice activity detection component would be required as the tool's input stage. Anyway, an approach like ours should never be applied directly for diagnostic purposes, it should rather raise a probable cause to initiate a diagnostic cascade (i. e., neurological, neuropaediatric, and genetic testing).

6. Conclusions and Outlook

Families with infants with RTT usually undergo periods of uncertainty with respect to their children's development until the diagnosis of RTT is made. In this study, we investigated the existence of potential acoustic signal level parameters in early vocalisations exploitable for facilitating an earlier identification of individuals with RTT in the context of the classification paradigm RTT versus typical development. We achieved promising recognition accuracies on the basis of 4 678 prelinguistic vocalisations of four infants later diagnosed with RTT and four TD infants when performing infant-dependent feature normalisation prior to classification. As far as we know, our study testified for the very first time that an objective approach to automatically identify infants with RTT in the first year of life based on vocalisation acoustics on signal level may be feasible and impact future earlier identification procedures.

Even though we built our study upon a considerable number of vocalisations per class, the overall number of infants per class was critically low. Therefore, we aim to expand our study by adding vocalisations of a high number of both infants later diagnosed with RTT and TD infants matched with respect to family language. We further aim to extend our age range of interest to the whole first year of life including the period that captures the first occurrences of melodic-modulated sounds/cooing in typical development [16, 17]. Based on a more extensive dataset, we shall treat different vocalisation types as well as vocalisations produced in different developmental stages in separate.

From a technological point of view, in future work a special focus should be put on the selection of acoustic features relevant or specific for RTT versus typical development or other neurodevelopmental disorders. A more detailed evaluation of different feature processing/feature normalisation/standardisation procedures should be carried out. Finally, the strengths and weaknesses of different classification approaches in context of this special area of application should be identified.

7. Acknowledgements

The authors acknowledge funding from the National Bank of Austria (OeNB; P16430), BioTechMed-Graz, the General Movements Trust, and the EU's H2020 Programme via RIA #688835 (DE-ENIGMA). Special thanks go to Jorge Luis Moye, Sergio Roccabado, Adriana Villarroel, and Claudia Zitta for their assistance in the vocalisation segmentation process. Moreover, we want to thank Katrin D. Bartl-Pokorny for consultancy in linguistic matters. The authors are grateful to all parents who provided us with home video material for scientific analysis.

8. References

[1] B. Hagberg, J. Aicardi, K. Dias, and O. Ramos, "A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases," *Annals of Neurology*, vol. 14, no. 4, pp. 471–479, 1983.

[2] J. L. Neul, W. E. Kaufmann, D. G. Glaze, J. Christodoulou, A. J. Clarke, N. Bahi-Buisson, H. Leonard, M. E. S. Bailey, N. C. Schanen, M. Zappella, A. Renieri, P. Huppke, and A. K. Percy, "Rett syndrome: Revised diagnostic criteria and nomenclature," *Annals of Neurology*, vol. 68, no. 6, pp. 944–950, 2010.

[3] C. L. Laurvick, N. de Klerk, C. Bower, J. Christodoulou, D. Ravine, C. Ellaway, S. Williamson, and H. Leonard, "Rett syndrome in Australia: A review of the epidemiology," *The Journal of Pediatrics*, vol. 148, no. 3, pp. 347–352, 2006.

[4] A. Rett, "Über ein zerebral-atrophisches Syndrom bei Hyperammonämie," *Wiener Medizinische Wochenschrift*, vol. 116, pp. 723–726, 1966.

[5] R. E. Amir, I. B. van den Veyver, M. Wan, C. Q. Tran, U. Francke, and H. Y. Zoghbi, "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2," *Nature Genetics*, vol. 23, no. 2, pp. 185–188, 1999.

[6] H. Cass, S. Reilly, L. Owen, A. Wisbeach, L. Weekes, V. Slonims, T. Wigram, and T. Charman, "Findings from a multidisciplinary clinical case series of females with Rett syndrome," *Developmental Medicine & Child Neurology*, vol. 45, no. 5, pp. 325–337, 2003.

[7] A. M. Kerr, H. L. Archer, J. C. Evans, R. J. Prescott, and F. Gibbon, "People with MECP2 mutation-positive Rett disorder who converse," *Journal of Intellectual Disability Research*, vol. 50, no. 5, pp. 386–394, 2006.

[8] J. L. Matson, J. C. Fodstad, and J. A. Boisjoli, "Nosology and diagnosis of Rett syndrome," *Research in Autism Spectrum Disorders*, vol. 2, no. 4, pp. 601–611, 2008.

[9] J. Sigafoos, D. Kagohara, L. van der Meer, V. A. Green, M. F. O'Reilly, G. E. Lancioni, R. Lang, M. Rispoli, and D. Zisimopoulos, "Communication assessment for individuals with Rett syndrome: A systematic review," *Research in Autism Spectrum Disorders*, vol. 5, no. 2, pp. 692–700, 2011.

[10] P. B. Marschik, C. Einspieler, and J. Sigafoos, "Contributing to the early detection of Rett syndrome: The potential role of auditory Gestalt perception," *Research in Developmental Disabilities*, vol. 33, no. 2, pp. 461–466, 2012.

[11] P. B. Marschik, G. Pini, K. D. Bartl-Pokorny, M. Duckworth, M. Gugatschka, R. Vollmann, M. Zappella, and C. Einspieler, "Early speech-language development in females with Rett syndrome: Focusing on the preserved speech variant," *Developmental Medicine & Child Neurology*, vol. 54, no. 5, pp. 451–456, 2012.

[12] K. D. Bartl-Pokorny, P. B. Marschik, J. Sigafoos, H. Tager-Flusberg, W. E. Kaufmann, T. Grossmann, and C. Einspieler, "Early socio-communicative forms and functions in typical Rett syndrome," *Research in Developmental Disabilities*, vol. 34, no. 10, pp. 3133–3138, 2013.

[13] P. B. Marschik, W. E. Kaufmann, J. Sigafoos, T. Wolin, D. Zhang, K. D. Bartl-Pokorny, G. Pini, M. Zappella, H. Tager-Flusberg, C. Einspieler, and M. V. Johnston, "Changing the perspective on early development of Rett syndrome," *Research in Developmental Disabilities*, vol. 34, no. 4, pp. 1236–1239, 2013.

[14] P. B. Marschik, R. Vollmann, K. D. Bartl-Pokorny, V. A. Green, L. van der Meer, T. Wolin, and C. Einspieler, "Developmental profile of speech-language and communicative functions in an individual with the Preserved Speech Variant of Rett syndrome," *Developmental Neurorehabilitation*, vol. 17, no. 4, pp. 284–290, 2014.

[15] D. K. Oller, R. E. Eilers, A. R. Neal, and A. B. Cobo-Lewis, "Late onset canonical babbling: A possible early marker of abnormal development," *American Journal on Mental Retardation*, vol. 103, no. 3, pp. 249–263, 1998.

[16] D. K. Oller, "The emergence of the sounds of speech in infancy," *Child Phonology*, vol. 1, pp. 93–112, 1980.

[17] ——, *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[18] R. E. Stark, "Stages of speech development in the first year of life," *Child Phonology*, vol. 1, pp. 73–92, 1980.

[19] ——, "Infant vocalization: A comprehensive view," *Infant Mental Health Journal*, vol. 2, no. 2, pp. 118–128, 1981.

[20] R. E. Stark, L. E. Bernstein, and M. E. Demorest, "Vocal communication in the first 18 months of life," *Journal of Speech, Language, and Hearing Research*, vol. 36, no. 3, pp. 548–558, 1993.

[21] J. L. Locke, *The Child's Path to Spoken Language*. Cambridge, Massachusetts: Harvard University Press, 1995.

[22] M. P. Lynch, D. K. Oller, M. L. Steffens, and E. H. Buder, "Phrasing in prelinguistic vocalizations," *Developmental Psychobiology*, vol. 28, no. 1, pp. 3–25, 1995.

[23] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE: The Munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM International Conference on Multimedia, MM 2010*. Florence, Italy: ACM, October 2010, pp. 1459–1462.

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

[25] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association, INTERSPEECH 2013*. Lyon, France: ISCA, August 2013, pp. 148–152.

[26] B. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval, E. Marchi, and Y. Zhang, "The INTERSPEECH 2014 computational paralinguistics challenge: Cognitive & physical load," in *Proceedings of the 15th Annual Conference of the International Speech Communication Association, INTERSPEECH 2014*. Singapore, Singapore: ISCA, September 2014, pp. 427–431.

[27] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling TEchnique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[29] G. T. Baranek, C. R. Barnett, E. M. Adams, N. A. Wolcott, L. R. Watson, and E. R. Crais, "Object play in infants with autism: Methodological issues in retrospective video analysis," *American Journal of Occupational Therapy*, vol. 59, no. 1, pp. 20–30, 2005.

[30] P. B. Marschik and C. Einspieler, "Methodological note: Video analysis of the early development of Rett syndrome – one method for many disciplines," *Developmental Neurorehabilitation*, vol. 14, no. 6, pp. 355–357, 2011.

[31] J. L. Adrien, P. Lenoir, J. Martineau, A. Perrot, L. Hameury, C. Larmande, and D. Sauvage, "Blind ratings of early symptoms of autism based upon family home movies," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 32, no. 3, pp. 617–626, 1993.

[32] R. Palomo, M. Belinchón, and S. Ozonoff, "Autism and family home movies: A comprehensive review," *Journal of Developmental & Behavioral Pediatrics*, vol. 27, no. 2, pp. 59–68, 2006.

[33] C. Saint-Georges, R. S. Cassel, D. Cohen, M. Chetouani, M.-C. Laznik, S. Maestro, and F. Muratori, "What studies of family home movies can teach us about autistic infants: A literature review," *Research in Autism Spectrum Disorders*, vol. 4, no. 3, pp. 355–366, 2010.