

Personalized Modeling of Real-World Vocalizations from Nonverbal Individuals

Jaya Narain**
 Kristina T. Johnson**
 jnarain@mit.edu
 ktj@mit.edu

Massachusetts Institute of Technology
 Cambridge, Massachusetts, USA

Craig Ferguson
 Amanda O'Brien
 craigsonoffergus@gmail.com
 amobrien@mit.edu

Massachusetts Institute of Technology
 Cambridge, Massachusetts, USA

Tanya Talkar
 Tanya.Talkar@ll.mit.edu
 MIT Lincoln Laboratory
 Lexington, Massachusetts, USA

Yue Zhang Weninger
 Peter Wofford
 yuefw@mit.edu
 peterwof@mit.edu

Massachusetts Institute of Technology
 Cambridge, Massachusetts, USA

Thomas Quatieri
 quatieri@ll.mit.edu
 MIT Lincoln Laboratory
 Lexington, Massachusetts, USA

Rosalind Picard
 Pattie Maes
 picard@media.mit.edu
 pattie@media.mit.edu
 Massachusetts Institute of Technology
 Cambridge, Massachusetts, USA

ABSTRACT

Nonverbal vocalizations contain important affective and communicative information, especially for those who do not use traditional speech, including individuals who have autism and are non- or minimally verbal (nv/mv). Although these vocalizations are often understood by those who know them well, they can be challenging to understand for the community-at-large. This work presents (1) a methodology for collecting spontaneous vocalizations from nv/mv individuals in natural environments, with no researcher present, and personalized in-the-moment labels from a family member; (2) speaker-dependent classification of these real-world sounds for three nv/mv individuals; and (3) an interactive application to translate the nonverbal vocalizations in real time. Using support-vector machine and random forest models, we achieved speaker-dependent unweighted average recalls (UARs) of 0.75, 0.53, and 0.79 for the three individuals, respectively, with each model discriminating between 5 nonverbal vocalization classes. We also present first results for real-time binary classification of positive- and negative-affect nonverbal vocalizations, trained using a commercial wearable microphone and tested in real time using a smartphone. This work informs personalized machine learning methods for non-traditional communicators and advances real-world interactive augmentative technology for an underserved population.

CCS CONCEPTS

• **Human-centered computing** → *Accessibility technologies*; • **Computing methodologies** → *Speech recognition*; Machine learning; Feature selection.

**Joint first authors; Both authors contributed equally to this research.



This work is licensed under a Creative Commons Attribution International 4.0 License.

ICMI '20, October 25–29, 2020, Virtual event, Netherlands
 © 2020 Copyright held by the owner/author(s).
 ACM ISBN 978-1-4503-7581-8/20/10.
<https://doi.org/10.1145/3382507.3418854>

KEYWORDS

nonverbal communication; paralinguistics; affect detection; autism; human-computer interaction; machine learning

ACM Reference Format:

Jaya Narain*, Kristina T. Johnson*, Craig Ferguson, Amanda O'Brien, Tanya Talkar, Yue Zhang Weninger, Peter Wofford, Thomas Quatieri, Rosalind Picard, and Pattie Maes. 2020. Personalized Modeling of Real-World Vocalizations from Nonverbal Individuals. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20)*, October 25–29, 2020, Virtual event, Netherlands. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3382507.3418854>

1 INTRODUCTION

Nonverbal vocalizations, such as sighs, grunts, and monosyllabic sounds, are a critical and natural component of speech. Typically, these sounds are used to enhance the affective and social exchanges surrounding speech [14]. However, individuals who are non- or minimally verbal (nv/mv) often communicate vocally without using typical verbal speech, using zero or fewer than 20 words or word approximations, respectively. These individuals include approximately 30% of individuals with autism spectrum disorder (ASD), as well as persons with certain genetic disorders or neurological differences [27, 29]. The vocalizations from these individuals convey rich affective and communicative information for a range of functions, but they are highly idiosyncratic to the speaker, making interpretation of these utterances incomprehensible to unfamiliar listeners. In addition, these sounds are dependent on individual motivation, mood, and environment, requiring study in real-world environments and processing methods that account for the sparsity, noise, and heterogeneity of sounds. As such, limited research exists examining the diverse phenotypes of nv/mv vocal communication or promoting reciprocal vocal exchanges between nv/mv individuals and others [13].

In order to increase awareness, expand understanding, and enhance communication for and with nv/mv individuals, we must create a pipeline of new technology, data collection methodologies, and analysis techniques that are built around the targeted

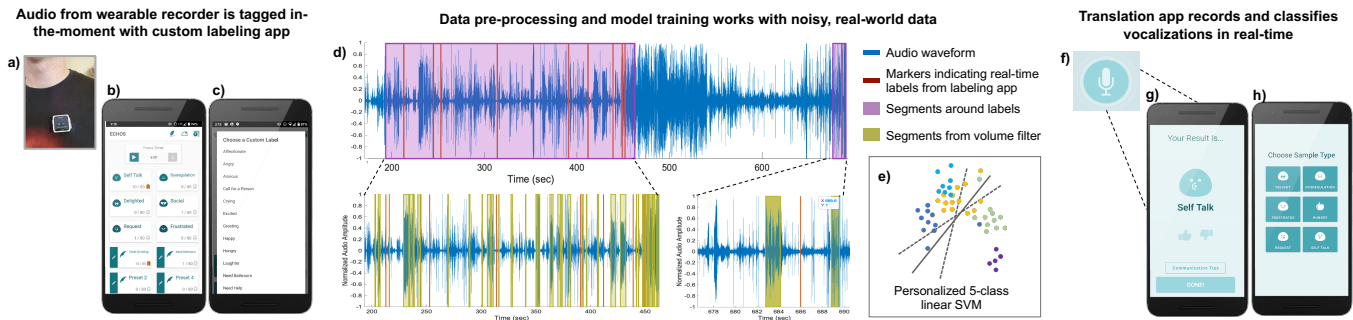


Figure 1: System overview: a) Wearable microphone and b-c) labeling app for in-situ labeling of nonverbal vocalizations using multimodal context. The timestamped labels are synced to a server and (d) aligned with audio segments isolated from the participant’s wearable microphone using a volume-based filter. (e) Personalized models are then trained for deployment in a real-time translation application. f) Pressing the microphone button records and sends a vocalization to a cloud-based server for real-time classification. g) In < 1 sec, the result is shown to the user, who can provide feedback. h) The user can also listen to personalized examples of each vocalization type to share with others to expand awareness and dyadic communication.

real-world impact. This paper presents the first results of such a system, including (1) a methodology for collecting and processing real-world audio that is labeled in real-time by a family member who knows the individual well, (2) multi-class speaker-dependent machine classification of nonverbal vocalizations by function, and (3) the design and implementation of a real-time translational tool to facilitate understanding of these unique vocalizations by the world-at-large.

2 RELATED WORK

To our knowledge, no other studies have acquired nonverbal vocalizations from child and adult nv/mv individuals using personalized labels in natural environments. Few language assessment protocols exist for preverbal individuals [28], and those that do almost universally rely on parent or clinician report [6, 13]. Oller et al. [20] characterized the acoustic content of vocalizations, including those without verbal content, in natural settings using a wearable device, but the focus was on diagnosing language delays in toddlers only and the vocalizations were unlabeled. Bacon et al. [4] created a large naturalistic dataset of toddler speech to compare the language development of 3 year olds with and without ASD by manually coding clinic-visit videos at the vocalization and word level, but this labor-intensive process has not been extended to older populations and uses only short lab-based interactions. None of these studies considered the function of the vocalizations.

However, infant cries, which lack verbal content, have been classified by function by both humans and machines [10, 16, 30]. There has also been significant work identifying emotions in verbal speech based on the acoustic properties generally [1, 21, 22, 25, 26], and for individuals with autism and verbal speech, specifically [17]. Additionally, nonverbal vocalizations that occur alongside traditional language have been studied anthropologically [2, 23, 24] and have been classified using machine learning methods with natural and acted vocalizations [3]. Researchers like Beukelmen and Miranda [5] have noted that nv/mv individuals use vocalizations to express emotions and communicate, but systematic study and tools to communicate these expressions remain undeveloped.

3 METHODOLOGY

3.1 Multimodal Data Collection

Participants were recruited through discussions of the project with the community. This study has been approved by an institutional review board (IRB), and recruitment is ongoing and open to any nv/mv individual. Throughout this text, the nv/mv individual will be referred to as the *participant*, while the person who knows the participant well enough to understand his/her vocalizations will be referred to as the *labeler*. Here, we present the results from three participants, aged 18 (P01), 23 (P02), and 8 (P03), all male with diagnoses of ASD. P01 and P03 are nonverbal, while P02 is minimally verbal with four consistent word approximations.

The audio samples were recorded in each participant’s natural environment, usually in and around his home. A small, wireless recorder (Sony IDC-TX800 recording in 16-bit, 44.1 kHz stereo) was attached to the participant’s clothing using strong magnets (Fig. 1a). Then, the labeler used a custom Android app (Fig. 1b) to label vocalizations the participant made by affective or communicative function in real-time. The app and labeling process were built through participatory design with the early users [11, 12]. Six labels were the same for all users – self-talk, delight, dysregulation, frustration, request, and social exchange – and were developed through conversations and testing with nv/mv individuals and families [19]. Each user also had the option of customizing up to 4 preset labels from a list of 25 affective or communicative functions (Fig. 1c). The participating families were encouraged to proceed with their usual activities during data collection and to upload the audio data from the recorder to IRB-approved researchers using a cloud-based service at their discretion. The study protocol supports fully remote participation, enabling access to an underrepresented, geographically distributed population, even amidst global health concerns like COVID-19.

3.2 Data Preprocessing

To minimize the data search space, the collected audio was first segmented around the locations of labels (see Fig. 1d). It was then

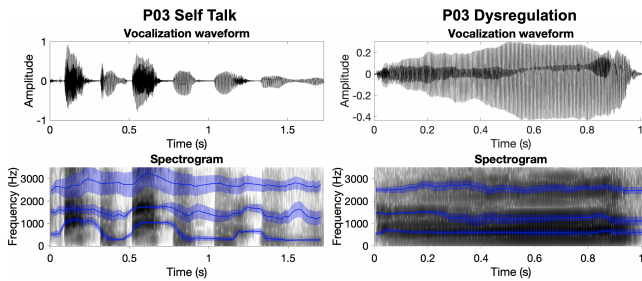


Figure 2: Sample waveforms (top) and spectrograms (bottom) for self-talk and dysregulation vocalizations from P03, showing tracked formant values (blue lines) with ± 3 dB bandwidth (blue shading), calculated using KARMA. Visual and audio examples from all participants and vocalization types are available in Supplements 1 and 2.

segmented by volume and associated with temporally adjacent labels, accounting for human delays in labeling. Each segment was verified manually to ensure that it contained a vocalization from the individual without excessive overlapping voices or noise. Some segments were further trimmed manually or using a voice-activity detector to remove noise before or after the vocalization. All vocalizations used in the analysis included samples from different recording days to reduce the likelihood of capturing distinctions purely related to the acoustic environment. The total number of vocalizations used in the present analysis, per function and participant, is given in Figure 3d.

3.3 Vocal Analysis

Post-hoc exploratory single-factor analysis of variance (ANOVA) tests were performed within participants and across vocalization types. Means and variances of the pitch, first, second, and third formants, as well as the longest sustained value and duration for the first and second formants in each vocalization segment, were examined as dependent variables. These formant values were calculated using KARMA [18], a Kalman-based autoregressive model for formant tracking (see Fig. 2). These values were used to inform a small custom feature set, as well as to expand the study of the underlying acoustical characteristics and phonological development of non-speech vocalizations from nv/mv individuals.

3.4 Feature Selection and Modeling

Models were trained and evaluated using four feature sets: the ComParE 2013 feature set [9] (dimension 6373), features selected from the ComParE set using Best First feature selection with greedy hill-climbing and backtracking [7] (participant-specific dimensions), the extended Geneva Minimalist Acoustic Parameter Set (eGeMAPS) [8] (dim. 88), and a custom feature set based on observed and hypothesized trends in the data (dim. 43). The custom feature set included statistics for each vocalization related to power, pitch, mel frequency cepstral coefficients (MFCCs), and formants, calculated using KARMA [18]. Formants and cepstral features were calculated using a Hamming window with a 20 ms frame and 0.5 overlap. The complete list of custom features is available in the supplement.

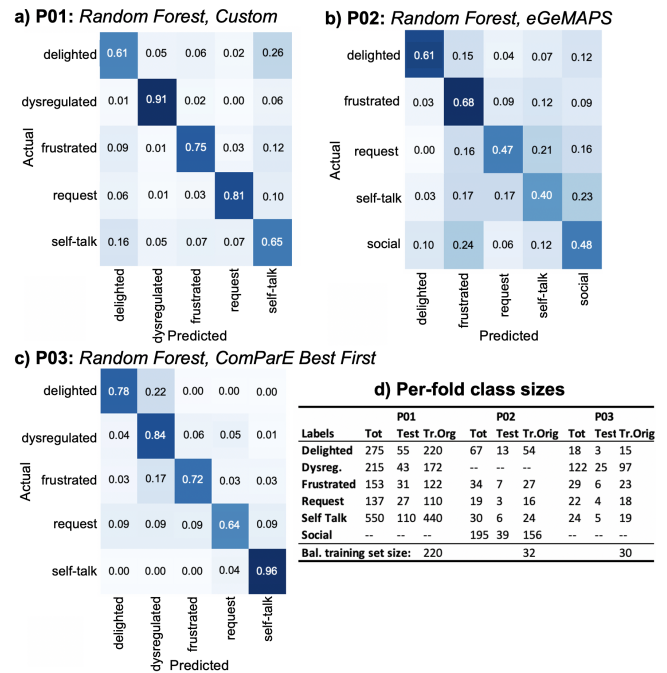


Figure 3: (a-c) Confusion matrices for the model and feature set with the highest UAR for each participant. (d) Table showing the total number of vocalization segments in the original dataset (Tot); the per-fold number of segments in the test set (Test), original training set (Tr.Orig), and balanced training subset after downsampling and/or synthetic data augmentation (Bal. training set size)

Five-way classifiers were trained and evaluated for each participant using nested 5-fold cross validation. For the Best First feature sets, features were selected from the full ComParE set using all the training data for the fold. For each SVM model, a balanced subset of the training fold was used to select an optimal regularization parameter via 5-fold cross validation. The training set for each fold was created by either downsampling randomly or upsampling using the Synthetic Minority Oversampling Technique (SMOTE) [15] to create balanced sets twice the smallest original training class size. Figure 3d lists the original number of samples for each class per participant, as well as the original and final training set sizes.

4 RESULTS & DISCUSSION

Table 1 shows the results with the evaluated classifiers, a Random Forest with 100 trees and an SVM trained using stochastic gradient descent, with each feature set. The unweighted average recall (UAR) values of the evaluated models in this study were all well above chance (0.2 for a five-way classification task), indicating that personalized machine-based classification of non-speech vocalizations from nv/mv individuals is feasible. The comparable performance between the small, 43-feature custom set and the larger feature sets illustrates that a small number of features can effectively capture differences between vocalization functions. Combined with the presented vocal analysis, these results suggest that there are inherent

within-speaker distinctions between vocalizations of different functions from nv/mv individuals. The custom feature set was primarily composed of pitch- and formant-related features and MFCC values, suggesting that these features may be particularly salient.

The variation of model UAR between participants is likely due to differences in data quality and quantity between labelers, in addition to inherent differences in vocalization practices between participants. For example, the labeler for P03 prioritized short, frequent labels with start and stop times immediately surrounding a vocalization. The labelers for P01 and P02 tended to label longer sets of vocalizations with a single label designating the overall emotion or event, increasing the likelihood of mislabeled vocalizations over the longer time period. While the large amount of data for P01 likely offset the labeling ambiguity, the small sample size for P02 did not have this advantage and produced lower UARs overall.

Figure 3a-c shows the normalized confusion matrices for a maximally performing classifier for each participant. High diagonal values support the UAR findings, while moderate off-diagonal values are consistent with imbalanced class sizes and/or classification overlap between similar affective or communicative classes (e.g., delighted/self talk for P01 or social/request for P02).

Table 1: Calculated UAR for 5-way classification task

	P01	P02	P03
<i>SVM</i>			
ComParE All	0.74	0.47	0.69
ComParE Best First	0.70	0.48	0.75
Custom	0.67	0.43	0.61
eGeMAPS	0.64	0.44	0.62
<i>Random Forest</i>			
ComParE All	0.73	0.52	0.68
ComParE Best First	0.73	0.47	0.79
Custom	0.75	0.42	0.70
eGeMAPS	0.68	0.53	0.73

5 REAL-TIME TRANSLATION

The presented offline classification results are a step towards a real-world, real-time, user-specific AAC system that enhances dyadic interaction and understanding between nv/mv individuals and the world. To that end, we developed a smartphone app that records a vocalization snippet and employs a personalized, pre-trained model to classify that individual’s vocalizations in real-time (see Fig. 1f-h). By pressing the microphone button (Fig. 1f), the app saves the recording, along with 2 seconds of sound prior to the button press to account for human delay. That 2-second buffer is continuously overwritten to preserve privacy. The phone then sends the audio to a server, which applies the personalized model and sends its prediction back to the app’s results screen (Fig. 1g). In addition, the “Sample Sounds” screen (Fig. 1h) allows users to play recordings of various vocalization classes from that nv/mv individual. This feature was included to introduce others to the meaning and range of vocalizations that the individual expresses and to encourage them to interpret these sounds as dialogue.

The app was pilot tested with P03 and his family. Audio data were recorded via the app with a Samsung Galaxy S10 smartphone in noisy home and outdoor environments and were not pre-processed. Using a pre-trained personalized SVM model, classification results were calculated and displayed to the user within one second of releasing the record button. Optionally, the user could confirm or correct the “translation” from the results page.

A binary classification model was trained to distinguish between two labeled states – self talk (positive affect) and dysregulation (negative affect). To more closely resemble audio recorded via a smartphone, each original audio segment from P03’s wearable recorder was passed through a 6th-order lowpass Butterworth filter with a cutoff frequency of 8 kHz, and the amplitude was multiplied by 0.22. These altered segments were combined with the original segments to create an augmented training set twice the original set size (n=48 self talk; n=244 dysreg). Best First features were selected from the ComParE feature set using [7]. Then the binary model was trained and validated using downsampled, balanced sets of 48 samples from the augmented dataset, and tested using a held-out, future test set of smartphone-recorded audio files. The resultant model had a 90.4% classification accuracy, with 14 self-talk true positives (TP), 1 self-talk false positive (FP), 5 dysregulation TP, and 1 dysregulation FP.

6 CONCLUSIONS AND FUTURE WORK

We have presented the first end-to-end system to collect, label, classify, and translate non-speech vocalizations from nv/mv individuals using personalized, real-world data. Using a remote data collection methodology that combines unobtrusive wearable audio recordings with in-situ smartphone labels from a family member, we conducted case studies with three nv/mv individuals. Five-way speaker-dependent classification produced UAR values of 0.75, 0.53, and 0.79 for P01, P02, and P03, respectively, and exploratory analysis revealed distinct vocal characteristics based on vocalization functions. To evaluate the potential of the system in-the-wild with one family, we trained and deployed a binary classifier trained using wearable recorder data and tested it with noisy smartphone data that accurately distinguished between affective vocalizations.

In future work, we will examine inter- and intra-person temporal-spectral and phonetic characteristics to bolster nv/mv communication research and to assess commonalities or differences that might improve classification performance. We will also explore the use of end-to-end-learning and semi-supervised learning with unlabeled data, develop a robust 5-way classifier for the translational mobile application, and expand data collection and recruitment. In order to improve data quality and labeling consistency across all participants, we have created a website and instructional videos to better communicate data collection procedures remotely and accessibly. Newly enrolled participants now receive a remote-accessible kit of materials with video and text tutorials, built through iterative feedback with these early participants, to ensure that any individual or family from this underrepresented, geographically distributed population can successfully join this work. Our vision is that real-time feedback will create new and strengthen existing communication channels, increasing the agency of the communicator and empowering everyone to share more.

REFERENCES

- [1] Mohammed Abdelwahab and Carlos Busso. 2015. Supervised Domain Adaptation for Emotion Recognition from Speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5058–5062.
- [2] Andrey Anikin and César F Lima. 2017. Perceptual and acoustic differences between authentic and acted nonverbal emotional vocalizations. *The Quarterly Journal of Experimental Psychology* (2017), 1–21.
- [3] Andrey Anikin and Tomas Persson. 2017. Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods* 49, 2 (2017), 758–771.
- [4] Elizabeth C Bacon, Suzanna Osuna, Eric Courchesne, and Karen Pierce. 2019. Naturalistic language sampling to characterize the language abilities of 3-year-olds with autism spectrum disorder. *Autism* 23, 3 (2019), 699–712.
- [5] David R. Beukelman and Patt Mirenda. 2013. *Augmentative and Alternative Communication: Supporting Children and Adults with Complex Communication Needs* (4th ed.). Paul H. Brookes Publishing.
- [6] Nancy C Brady, Kandace Fleming, Kathy Thiemann-Bourque, Lesley Olswang, Patricia Dowden, Muriel D Saunders, and Janet Marquis. 2012. Development of the communication complexity scale. *American Journal of Speech-Language Pathology* (2012).
- [7] Frank Eibe, Mark A. Hall, and Ian H. Witten. 2016. *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"* (4 ed.). Morgan Kaufmann.
- [8] Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. 2015. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing* 7, 2 (2015), 190–202.
- [9] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. 2013. Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor. In *Proceedings of the 21st ACM International Conference on Multimedia*. 835–838.
- [10] Tanja Fuhr, Henning Reetz, and Carla Wegener. 2015. Comparison of supervised-learning models for infant cry classification/vergleich von klassifikationsmodellen zur säuglingsschreianalyse. *International Journal of Health Professions* 2, 1 (2015), 4–15.
- [11] Kristina T Johnson*, Jaya Narain*, Craig Ferguson, Rosalind Picard, and Pattie Maes. 2020. The ECHOS Platform to Enhance Communication for Nonverbal Children with Autism: A Case Study. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [12] Kristina T Johnson*, Jaya Narain*, Rosalind W Picard, and Pattie Maes. 2020. Augmenting Natural Communication in Nonverbal Individuals with Autism. In *International Society of Autism Research (INSAR) 2020 Meeting*. INSAR.
- [13] Connie Kasari, Nancy Brady, Catherine Lord, and Helen Tager-Flusberg. 2013. Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research* 6, 6 (2013), 479–493.
- [14] Mark L Knapp. 1980. *Essentials of Nonverbal Communication*. Harcourt School.
- [15] Guillaume Lemaitre, Fernando Nogueira, and Christos K. Aridas. 2017. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research* 18, 17 (2017), 1–5.
- [16] Lichuan Liu, Wei Li, Xianwen Wu, and Benjamin X Zhou. 2019. Infant cry language analysis and recognition: an experimental approach. *IEEE/CAA Journal of Automatica Sinica* 6, 3 (2019), 778–788.
- [17] Erik Marchi, Björn Schuller, Anton Batliner, Shimrit Fridenzon, Shahar Tal, and Ofer Golan. 2012. 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*.
- [18] Daryush D Mehta, Daniel Rudoy, and Patrick J Wolfe. 2012. Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking. *The Journal of the Acoustical Society of America* 132, 3 (2012), 1732–1746.
- [19] Jaya Narain*, Kristina T Johnson*, Amanda O'Brien, Peter Wofford, Pattie Maes, and Rosalind Picard. 2020. Nonverbal Vocalizations as Speech: Characterizing Natural-Environment Audio from Nonverbal Individuals with Autism. In *Laughter and Other Non-Verbal Vocalisations Workshop 2020*.
- [20] D Kimbrough Oller, P Niyogi, S Gray, Jeffrey A Richards, Jill Gilkerson, Daoyi Xu, Umit Yapanel, and Steven F Warren. 2010. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences* 107, 30 (2010), 13354–13359.
- [21] Anda Ouyang, Ting Dang, Vidhyasaharan Sethu, and Eliathamby Ambikairajah. 2019. Speech Based Emotion Prediction: Can a Linear Model Work? *Proc. INTERSPEECH 2019* (2019), 2813–2817.
- [22] Marc D Pell, Silke Paulmann, Chinar Dara, Areej Alasser, and Sonja A Kotz. 2009. Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics* 37, 4 (2009), 417–435.
- [23] Disa A Sauter, Frank Eisner, Andrew J Calder, and Sophie K Scott. 2010. Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology* 63, 11 (2010), 2251–2272.
- [24] Disa A Sauter, Frank Eisner, Paul Ekman, and Sophie K Scott. 2010. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences* 107, 6 (2010), 2408–2412.
- [25] Maximilian Schmitt, Nicholas Cummins, and Björn Schuller. 2019. Continuous Emotion Recognition in Speech – Do We Need Recurrence? *Proc. INTERSPEECH 2019* 34, 93 (2019), 12.
- [26] Björn W Schuller. 2018. Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends. *Commun. ACM* 61, 5 (2018), 90–99.
- [27] Helen Tager-Flusberg and Connie Kasari. 2013. Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. *Autism Research* 6, 6 (2013), 468–478.
- [28] Helen Tager-Flusberg, Sally Rogers, Judith Cooper, Rebecca Landa, Catherine Lord, Rhea Paul, Mabel Rice, Carol Stoel-Gammon, Amy Wetherby, and Paul Yoder. 2009. Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research* (2009).
- [29] Irena Vrečar, Josie Innes, Elizabeth A Jones, Helen Kingston, William Reardon, Bronwyn Kerr, Jill Clayton-Smith, and Sofia Douzgou. 2017. Further clinical delineation of the MEF2C haploinsufficiency syndrome: report on new cases and literature review of severe neurodevelopmental disorders presenting with seizures, absent speech, and involuntary movements. *Journal of Pediatric Genetics* 6, 03 (2017), 129–141.
- [30] O Wasz-Höckert, TJ Partanen, V Vuorenkoski, K Michelsson, and E Valanne. 1964. The identification of some specific meanings in infant vocalization. *Experientia* 20, 3 (1964), 154–154.