Prosody Features Characterization of Autism Speech for Automated Detection and Classification

Abhijit Mohanta[1], Prerana Mukherjee[2]
[1,2]*Indian Institute of Information Technology, Sri City, Chittoor, Andhra Pradesh, India*
[1]abhijit.mohanta@iiits.in, [2]prerana.m@iiits.in

***Abstract*—The verbal children affected with *autism spectrum disorder* (ASD) often shows some notable acoustic patterns. This paper represents the classification of autism speech, i.e., the speech signal of children affected with ASD. In addition, this work specifically aims to classify the speech signals of non-native Indo English speakers (children) affected with ASD. Previous studies, however, have focused only on native English speakers. Hence, for this study purpose a speech signal dataset of ASD children and a speech signal dataset of normal children were recorded in English, and all the children selected for the data collection were non-native Indo English speakers. Here, for the ASD and the normal children, the acoustic features explored for classification are namely, fundamental frequency (F0), strength of excitation (SoE), formants frequencies (F1 to F5), dominant frequencies (FD1, FD2), signal energy (E), zero-crossing rate (ZCR), mel-frequency cepstral coefficients (MFCC), and linear prediction cepstrum coefficients (LPCC). In addition, VGGish audio features are also explored here. Further, these feature sets are classified by utilizing different classifiers. The KNN classifier model achieves the highest 96.5% accuracy with respect to other baseline models explored here.**

***Index Terms*—ASD children, acoustic features, MFCC, LPCC, SVM**

I. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder which involves communication deficits, social interaction impairments, and hyperfocus or reduced behavioral flexibility [1]. According to [2], 1 in 68 children affected with autism was reported in 2014. In fact, there are no fixed biological criteria to describe autism, also its specific characteristics and underlying mechanisms are still not decipherable [3]. Nevertheless, only a few studies have been carried out on the speech signal of ASD children, especially on the speech signal of non-native English speakers (children) affected with ASD.

The verbal individuals with ASD speak with atypical acoustic patterns, and hence they face difficulty in social interactions [4]. Also, the disturbances of prosody one of the most significant problems among verbal individuals with ASD [3]. Even, sometimes children with ASD show a notable spoken language delay and repetitive language.

Previous studies on children with ASD were mostly based on either speech prosody or unusual suprasegmental features of speech production [3]. Some of the most significant pitch based analyses of ASD children were reported in [5], where authors have reported higher mean pitch for ASD children than the normal children. Whereas, in [6], the authors have reported the opposite result, i.e., normal children indicate higher mean pitch than the ASD children. Besides, in the case of intensity based analysis reported in [6], authors have reported higher mean intensity for ASD children as compared with the normal children. On the other hand, in [7] authors have reported the opposite results. Likewise, based on duration (word duration, syllable duration, utterance duration, etc.), speech rate, voice patterns, etc., some significant results have been reported in [8].

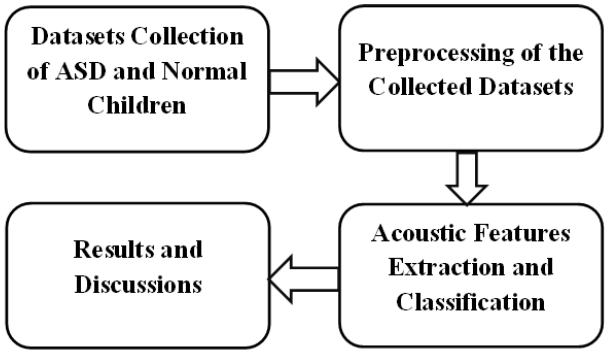
Datasets Collection of ASD and Normal Children
Preprocessing of the Collected Datasets
Acoustic Features Extraction and Classification
Results and Discussions

Fig. 1: Proposed plan's *block diagram.*

The aim of this paper is to differentiate the ASD children from the normal children. Differences are measured in terms of automatic classification by utilizing several classifiers, described in details in the later sections. Also, the differences are made in the context of statistical measurement of the acoustic features. Two speech signal datasets[1] are collected for this research purpose. One dataset contains speech samples of ASD children and another dataset contains the speech samples of normal children. Datasets details are given in Section 2.

This study has high significance because of several reasons. Firstly, in this study, the collected datasets are different in several ways with respect to the datasets collected previously. In our collected datasets all the speakers are non-native (Indian accent) English speakers. Whereas, in the earlier studies like [6], authors have only considered the native English speakers. It is important to study the speech signals of non-native Indo English speakers with ASD. Because, non-native Indo English speakers pronounce the English words letter wise, whereas native English speakers pronounce the English words phoneme

[1] Our collected *ASD affected* and *normal* children's speech signal datasets are available on request basis.

TABLE I: Datasets details of the (b1) ASD and (b2) normal children, where (a) represents several attributes and (b) represents statistical measurements of the ASD and normal speech signal datasets

(a) Characteristic	(b) Statistics	
	(b1) ASD	(b2) Normal
Number of Children	13	20
Age (Years)	03 to 09	03 to 09
Native Languages	Tamil and Telugu	Tamil and Telugu
English Reading Skill	Beginner level	Beginner level
Datasets Duration	9350 Seconds	12000 Seconds

wise. This reason makes the differences in the acoustic features of the speech signals of non-native Indian English speakers and native English speakers. Also, our datasets are recorded by asking the ASD and normal children to pronounce a same set of words over all the recording sessions. But, in previous studies authors have collected datasets mostly from spontaneous productions [3], [5], social interactions [8], and constrained production of ASD and normal children. Secondly, many robust speech features, especially dominant frequencies (FD1, FD2), strength of excitation (SoE) and fifth formant frequency (F5) have not been explored in previous studies. But, we got some significant results using those acoustic features, details in the results section. Finally, ASD current diagnostic criteria like DSM–IV, do not include any atypical vocalizations conditions [3], hence this study results can be used as acoustic markers for ASD.

This study is primarily divided into four major steps, as depicted in Figure 1. A brief overview of this current study is as follows. Firstly, two speech signal datasets were recorded, by recording the sound files of the ASD children and normal children. Secondly, in the preprocessing step, signal noise and unwanted signal parts were removed, and the speech signal files were arranged in two different databases for ASD and normal children. Thirdly, several speech signal processing methods were applied on the collected datasets to extract the selected speech production features. Also, several classifiers were used to classify the ASD and normal children in terms of their certain acoustic features. Lastly, results were made by differentiating between the ASD and normal children in terms of their speech production features and automatic classification results.

The arrangement of the rest of the paper is as follows. Details about the two collected datasets are given in Section 2. The speech signal processing methods and classifiers used in this study are discussed in Section 3. Then, Section 4 contains results and analyses of results. Lastly, Section 6 contains conclusion, also the future work scope on this topic.

II. Datasets Collection and Preprocessing

A. Datasets Collection

A speech signal dataset of ASD affected children and another speech signal dataset of normal children were collected for our study purpose. The details of both the datasets are tabulated in Table I. Apart from the information given in Table I, the children aged below 3 years were not considered because here only the verbal children were taken into consideration. Typically the children less than 36 months are non-verbal. The datasets were recorded in English. The main reason behind this was the group of children considered for data collection did not have the same native language. A group of the children had Tamil as their native language and other group had Telugu as their native language. So, we decided to record in English, because all of the selected children had a beginner-level English reading skill. Before the data collection took place it was made sure by a certified doctor the ASD children group considered hare for data collection were diagnosed with ASD, and they were under treatment during the period of data collection. Also, it was made sure by the doctor that all the ASD children considered here had some problems with their speaking during the entire period of data collection. On the other hand, the normal children considered here for data collection had no such issues and were living a normal and healthy life.

Speech samples of both the ASD and normal children group were collected weekly once or twice. We did not collect datasets in daily basis or in a single day because it may lose the children's interest to perform well. Also, it may lose the children's neutral emotional state during the data collection period. A noise-free closed room was selected for data samples recording. The room did not have any objects to distract the speaker. Also, we selected the recording place in such a position in the room that there is no sound echo related problem during recording sessions.

Furthermore, the process of the recording was as follows. The children were instructed to pronounce a set of 25 specially selected English words and numbers, shown to them alone with respective pictures. All the selected words and numbers were either in consonant-vowel-vowel-consonant (CVVC) or consonant-vowel-consonant (CVC) word format. The selected set of words contains the name of flowers, vegetables, animals, and English numbers. The set of words was written in English and shown to the children on a laptop and asked them to pronounce the word. Then, we kept changing the words with respective pictures one by one, and the speaker named the object (in English) which was displayed as a picture along with the picture name i.e., the English word. The set of 25 words along with pictures consist of 5 words from each English vowel. So, in each recording session, each child pronounced a set of 25 words, and we took such two sessions for each child, each day. Besides, we recorded the data samples using a Roland R-26 digital audio recorder with the sampling frequency as 48 KHz and in stereo recording mode. All the data samples were saved in .wav file format.

B. Preprocessing

The preprocessing was done two steps. First the unwanted parts of the recorded speech files were removed manually by using the wavesurfer tool. Second the signal noise was

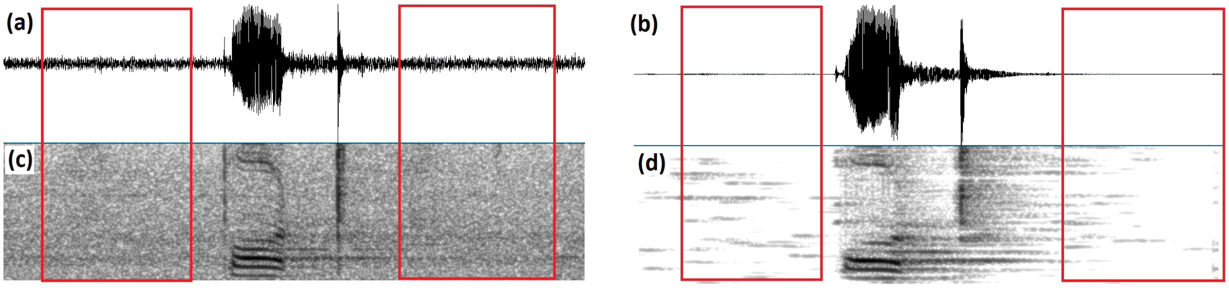

Fig. 2: Waveform and spectrogram of a same word (/tom/) before and after removing the signal noise. Here, (a) and (b) represent two waveforms before and after removing the signal noise, respectively, and (c) and (d) represent spectrograms of those two signals (in (a) and (b)) before and after removing the signal noise, respectively. The robustness of the noise removal method MMSE_VAD for *ASD* children can be visualize by observing the differences in the red colored highlighted parts of (c) and (d). The Y-axis varies from 0 Hz to 5000 Hz in the case of the spectrograms.

TABLE II: Quantitative Measurements (QM) of several noise cancellation algorithms for [A] *ASD* and [B] *Normal* children. Here, SS stands for spectral subtraction, MMSE stands for minimum mean squire error and LMV stands for Log MMSE voice activity detector (Log MMSE_VAD) based noise removal algorithm

QM	**[A] ASD**			**[B] Normal**		
	SS	**MMSE**	**LMV**	**SS**	**MMSE**	**LMV**
SNRseg	1.39	1.08	1	−4.27	−4.79	−4.8
PESQ	3.12	3.15	3.1	3.28	3.17	3.26

removed. In this study, three various state-of-the-art methods for speech signal's noise removal namely, spectral subtraction (SS), minimum mean square error (MMSE), and Log MMSE with voice activity detection (VAD) based algorithms were tried to remove the signal noise [9]. Then the performance of this three algorithms were measured by perceptual evaluation of speech quality (PESQ) and segmental SNR (SNRseg) quantitative measurements [10]. The SNRseg and PESQ gave a contradicting result, as tabulated in Table II. So to overcome that SNRseg results are ignored, because PESQ results are more promising than SNRseg in the case of speech signals. According to the PESQ results tabulated in Table II, it is clear that for ASD children Log MMSE_VAD method gives the best performance, and for normal children MMSE method gives the best result. Hence, to remove noise from ASD dataset the Log MMSE_VAD method was chosen, and to remove noise from normal children's speech dataset MMSE methods was used. Besides, in Figure 1, it is graphically represented a same signal before and after removing the signal noise. The performance of the selected noise removal method for ASD children i.e., MMSE_VAD can be visualize from Figure 2.

III. Speech Production Features and Autism Classification

A. Speech Production Features

The differences between the ASD and normal children are made in terms of their speech production characteristics and through the automatic classification using several classifiers. In this section both the acoustic features and classifiers are described. The source characteristics of speech production system are examined through fundamental frequency i.e., F0 and strength of excitation (SoE). The vocal tract (VT) filter characteristics are examined through first five formants frequencies i.e., F1 to F5 and dominant frequencies (FD1, FD2). Ten the source-filter combined characteristics are examined through zero-crossing rate (ZCR), signal energy (E). Here, for the statistical analyses purpose the mean (μ) and standard deviation (SD) values of all these above mentioned acoustic features are computed. The mean and SD values are computed by considering all the recording sessions for each speaker. The mean and SD values are calculated for each speaker by considering all the calculated values of a particular speech feature.

The F0 was derived using zero-frequency filtering (ZFF) method [11] method with the sampling frequency taken as 48 KHz. According to [11] the ZFF signal is defined as:

$$y[n] = y_2[n] - \frac{1}{2N+1} \sum_{m=-N}^{N} y_2[n+m] \tag{1}$$

where, window length is 2N+1 (in sample number), and $y_2[n]$ is the output of 2nd zero-frequency resonator followed by $y_1[n]$. The resultant signal is called the ZFF signal. The positive giving zero crossings of equation (1) gives the glottal closure instants, and these glottal closure instants are used to compute the F0 [11]. Next, the ZFF signal's slope (y[n]) around the glottal closure instants provides a measure of the SoE [11].

In the case of VT filter characteristics, the first five formants i.e., F1 to F5 are computed using linear prediction (LP) spectrum [12] with the sampling frequency (Fs) was taken as 10 KHz and LP order as 10. In addition, the first two dominant peak frequencies i.e., FD1 and FD2 are also derived

TABLE III: Classification results using different (a) classifiers with three different (b) cross validations (CV), along with classification (c) accuracy (Acc) in %, (d) Sensitivity (Sen), (e) specificity (Spe), (f) precision (Pre), (g) F1-score (F1-s) and (h) area under the ROC curve i.e., AUC

(a) Classifiers	**(b) CV**	**(c) Acc**	**(d) Sen**	**(e) Spe**	**(f) Pre**	**(g) F1-s**	**(h) AUC**
SVM (CK)	5-fold	92.9	0.94	0.92	0.92	0.93	0.93
KNN	5-fold	93.7	0.94	0.94	0.93	0.94	0.98
LD	5-fold	92.7	0.90	0.96	0.96	0.93	0.97
DT	5-fold	77.6	0.78	0.77	0.77	0.77	0.78
SVM (QK)	8-fold	92.4	0.93	0.92	0.91	0.92	0.97
KNN	8-fold	96.0	0.97	0.95	0.94	0.96	0.96
QD	8-fold	91.9	0.90	0.94	0.94	0.92	0.97
LR	8-fold	87.2	0.88	0.86	0.86	0.87	0.93
SVM (MGK)	10-fold	93.7	0.94	0.94	0.93	0.94	0.98
KNN	*10-fold*	*96.5*	*0.97*	*0.96*	*0.96*	*0.96*	*0.96*
LR	10-fold	88.4	0.88	0.89	0.89	0.88	0.95

using LP analysis with the LP order 5 and Fs 10 KHz. Then, the LP spectrum will have a maximum of two peaks, and the corresponding frequencies of these two peaks are denoted as FD1 and FD2, respectively [13]. The FD1 and FD2 give an idea of the energy concentration in the spectrum [13].

The source-filter combined characteristics E is computed using 25 ms frame size and 10 ms frame shift. According to [14] the E is defined in the context of discrete-time signal as

$$E_w = \sum_{n=-w/2}^{w/2} |x[n]|^2 \tag{2}$$

where, w represents window length. Next, the ZCR definition as provided in [15] is:

$$Z_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]| \, w(n-m) \tag{3}$$

where, $sgn[x(n)] = \begin{cases} 1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases}$ and

$w(n) = \begin{cases} \frac{1}{2N} \, for, \, 0 \leq n \leq N-1 \\ 0 \, for, \, otherwise \end{cases}$

Furthermore, only for the classification purpose, in order to get higher accuracy we have explored two more features i.e., MFCC [16] and LPCC [16]. In addition, only the first 13 MFCCs and 12 LPCCs coefficients are taken into consideration, because first 13 and 12 are the most significant MFCC and LPCC coefficients, respectively. The MFCCs are derived using 25 ms frame size and 10 ms of frame shift, and LPCCs are derived by taking the Fs as 10 KHz.

Also, state-of-the-art audio feature VGGish [17] is also explored in this study for a better classification accuracy. The VGGish is a 128 dimensional audio feature extractor model which is trained using YouTube-8M dataset.

B. Classifiers Design

In order to classify the ASD and normal children based on their speech production features, several classifiers are explored in this study. These classifiers are support vector machine (SVM) [18], K-nearest neighbors (KNN) [19], linear discriminant (LD) [20], quadratic discriminant (QD) [21] decision tree (DT) [22], and logistic regression (LR) [23]. In the case of SVM, the cubic kernel (CK), quadratic kernel (QK), and medium Gaussian kernel (MGK) are explored. Besides, in the case of KNN, 10 neighbors are used.

TABLE IV: The [A] mean (μ) and [B] SD (σ) values of acoustic (a) features of *ASD affected* and *Normal* children; (b) and (d) represents the acoustic features values for *ASD* children, and (c) and (e) represents the acoustic features values for *Normal* children

(a) Features	**[A] Mean**		**[B] SD**	
	(b) ASD	**(c) Normal**	**(d) ASD**	**(e) Normal**
F0	313	293	48	39
SoE	0.278	0.295	0.054	0.051
E	0.006	0.004	0.006	0.003
ZCR	0.087	0.108	0.021	0.022
F1	606	632	62	65
F2	1520	1483	104	75
F3	2636	2590	111	67
F4	3710	3671	89	57
F5	4373	4361	36	36
FD1	1088	1078	154	118
FD2	3045	3062	141	129

The performances of all the classifiers are observed through 5-fold, 8-fold and 10-fold cross-validations. Then for each cross-validation, only some of the most accurate (in terms of classification accuracy) classifiers models are taken into consideration. Also, the classifiers performances are measured through F1-score, ROC curve, and some other parameters given in Table III.

IV. Results and Observations

The observed results indicate that ASD children have higher μ_{F0} and μ_E values than the normal children. This result infers that ASD children have higher vocal fold vibration rate than the normal children. It also implies that ASD children put more vocalization effort than the normal children. Next, in the case of μ_{SoE}, the ASD children have lower value than the normal children. It implies that during the vibration of the vocal folds the strength of impulse-like excitation is lower for ASD children than the normal children. Next, the μ_{ZCR} has higher value for normal children than the ASD children. All these results are tabulated in Table IV.

Now in the case of VT filter features, μ_{F2}, μ_{F3}, μ_{F4}, and μ_{F5}, have higher values for ASD children than the normal children. This result infers that the pharyngeal-oral tract is shorter in length for ASD children than the normal children. Because, all the formants values are inversely proportional to the pharyngeal-oral tract length. Next, μ_{F1} has lower value for ASD children than the normal children. In terms of pharyngeal constriction, this result implies that ASD children have a lesser pharyngeal constriction than the normal children. Besides, μ_{FD1} have higher value for the ASD children and μ_{FD1} have

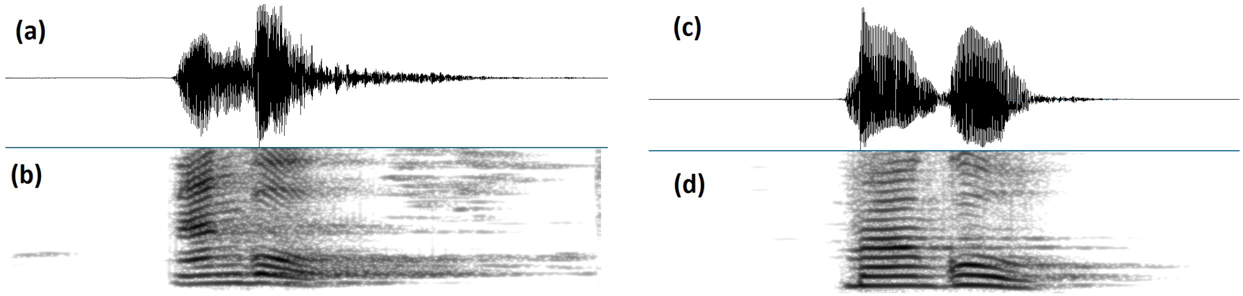

Fig. 3: Waveform and spectrogram of a same word (/mango/) pronounced by a child with *ASD* (a,b) and a *normal* child (c, d). The differences in *F0* and *Formants frequencies* values are clearly understandable between a ASD and normal child. The Y-axis varies from 0 Hz to 5000 Hz in the case of the spectrograms.

TABLE V: Comparison of our results with some of the significant previous studies: (a) represents authors and their paper reference, (b) represents classifiers name, and (c) represents classification accuracy in percentage (%)

(a) Authors	(b) Classifiers	(c) Accuracy (%)
Fusaroli et al., [24]	QD, linear regression	86.0
Oller et al., [25]	LD analysis	86.0
Santos et al., [26]	SVM	79.1
Kakihara et al., [27]	SVM	74.9
Santos et al., [26]	probabilistic neural network (PNN)	97.7
Proposed method	*KNN*	*96.5*

TABLE VI: *Classification accuracy using VGGish audio features.* Classification results using different (a) classifiers with three different (b) cross validations (CV), along with classification (c) accuracy (Acc) in %, (d) Sensitivity (Sen), (e) specificity (Spe), (f) precision (Pre), (g) F1-score (F1-s) and (h) area under the ROC curve i.e., AUC

(a) Classifiers	(b) CV	(c) Acc	(d) Sen	(e) Spe	(f) Pre	(g) F1-s	(h) AUC
SVM (QK)	5-fold	52.0	.51	0.52	0.42	0.46	0.53
SVM (CK)	5-fold	50.4	0.49	0.52	0.55	0.49	0.49
SVM (CK)	8-fold	50.7	0.49	0.52	0.58	0.53	0.53
SVM (QK)	10-fold	50.9	0.5	0.53	0.6	0.54	0.51

lower value for the ASD children as compared with the normal children. All these results can be analyzed from Table IV.

In the case of SD (σ), observations are as follows. The ASD children have higher σ_{F0}, σ_E, σ_{SoE}, σ_{F2}, σ_{F3}, σ_{F4}, σ_{FD1}, and σ_{FD2} values than the normal children, as tabulated in Table IV. In addition, in the case of σ_{FZCR} and σ_{F1}, ASD children have lower values than the normal children. But, the difference is non-significant. Lastly, as given in Table IV, σ_{F5} has the same values for the ASD and normal children.

The observations of the automatic classification results are as follows. The KNN classifier with 10 nearest neighbors (K value) and 10-fold cross validation gives the highest accuracy (96.5%), as tabulated in Table III. Also, in the context of all the three categories of cross-validations explored here i.e., 5-fold, 8-fold and 10-fold, KNN gives the highest accuracy in every category than other classifiers in those respective categories. Here our classification accuracy (96.5%) is significantly higher than most of the previous studies. Also, it is observed that classification using VGGish audio features gives lesser accuracy rate (52.0 %) as compared with classification using the handcrafted features. Finally, some of the comparisons with previous results are tabulated in tabulated in Table V.

V. Conclusion

This paper presents a novel technique to classify ASD children from normal children, in terms of their speech production features. An indigenous speech signal dataset of ASD children and another speech signal dataset of normal children are recorded for this work. The differences encoded in the acoustic features namely, F0, E, SoE, ZCR, first five formants frequencies, FD1, FD2, MFCC and LPCC features are utilized for classification. The classification accuracy is demonstrated by utilizing various classifiers. We validated through exhaustive experiments that there is a significant distinction between the ASD and normal children. We envisage that the robust results obtained in this work can be utilized as an acoustic biomarker to identify ASD from the speech signal at a very early age. Also, these robust results obtained from Indo English children with ASD can be compared with native English children with ASD, in future studies.

Acknowledgment

Authors are extremely grateful to Dr. N. P. Karthikeyen and DOAST Integrated Therapy Centre for Autism, Chennai, India, for providing us the opportunity to record ASD affected children's speech samples. The authors would also like to express their heartiest gratitude to Chinmaya Vidhyalaya, Sri City, Andhra Pradesh, India, for providing us the opportunity to record the normal children's speech samples. Most importantly, authors offer special thanks to each child participated in the data collection process.

References

[1] J. McCann and S. Peppé, "Prosody in autism spectrum disorders: a critical review," *International Journal of Language & Communication Disorders*, vol. 38, no. 4, pp. 325–350, 2003.

[2] Autism and D. D. M. N. S. Y. . P. Investigators, "Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2010," *Morbidity and Mortality Weekly Report: Surveillance Summaries*, vol. 63, no. 2, pp. 1–21, 2014.

[3] Y. S. Bonneh, Y. Levanon, O. Dean-Pardo, L. Lossos, and Y. Adini, "Abnormal speech spectrum and increased pitch variability in young autistic children," *Frontiers in human neuroscience*, vol. 4, p. 237, 2011.

[4] R. Fusaroli, A. Lambrechts, D. Bang, D. M. Bowler, and S. B. Gaigg, "Is voice a marker for autism spectrum disorder? a systematic review and meta-analysis," *Autism Research*, vol. 10, no. 3, pp. 384–407, 2017.

[5] M. G. Filipe, S. Frota, S. L. Castro, and S. G. Vicente, "Atypical prosody in asperger syndrome: Perceptual and acoustic measurements," *Journal of autism and developmental disorders*, vol. 44, no. 8, pp. 1972–1981, 2014.

[6] J. Quigley, S. McNally, and S. Lawson, "Prosodic patterns in interaction of low-risk and at-risk-of-autism spectrum disorders infants and their mothers at 12 and 18 months," *Language Learning and Development*, vol. 12, no. 3, pp. 295–310, 2016.

[7] L. A. Scharfstein, D. C. Beidel, V. K. Sims, and L. R. Finnell, "Social skills deficits and vocal characteristics of children with social phobia or asperger's disorder: A comparative study," *Journal of abnormal child psychology*, vol. 39, no. 6, pp. 865–875, 2011.

[8] J. Brisson, K. Martel, J. Serres, S. Sirois, and J.-L. Adrien, "Acoustic analysis of oral productions of infants later diagnosed with autism and their mother," *Infant mental health journal*, vol. 35, no. 3, pp. 285–295, 2014.

[9] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[10] Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," in *Ninth International Conference on Spoken Language Processing*, 2006.

[11] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1602–1613, 2008.

[12] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[13] V. K. Mittal and B. Yegnanarayana, "Analysis of production characteristics of laughter," *Computer Speech & Language*, vol. 30, no. 1, pp. 99–115, 2015.

[14] A. Rihaczek, "Signal energy distribution in time and frequency," *IEEE Transactions on information Theory*, vol. 14, no. 3, pp. 369–374, 1968.

[15] R. Bachu, S. Kopparthi, B. Adapa, and B. Barkana, "Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal," in *American Society for Engineering Education (ASEE) Zone ference Proceedings*, 2008, pp. 1–7.

[16] Y. Yujin, Z. Peihua, and Z. Qun, "Research of speaker recognition based on combination of lpcc and mfcc," in *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, vol. 3. IEEE, 2010, pp. 765–767.

[17] P. Guyot, "Simple cnn and vggish model for high-level sound categorization within the making sense of sounds challenge."

[18] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.

[19] B. V. Dasarathy, "Nearest neighbor (nn) norms: Nn pattern classification techniques," *IEEE Computer Society Tutorial*, 1991.

[20] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.

[21] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, "Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions," *Current applied physics*, vol. 11, no. 3, pp. 740–745, 2011.

[22] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142–147, 1977.

[23] R. Greiner, X. Su, B. Shen, and W. Zhou, "Structural extension to logistic regression: Discriminative parameter learning of belief net classifiers," *Machine Learning*, vol. 59, no. 3, pp. 297–322, 2005.

[24] R. Fusaroli, D. Bang, and E. Weed, "Non-linear analyses of speech and prosody in asperger's syndrome," in *International Meeting For Autism Research*, 2013.

[25] D. K. Oller, P. Niyogi, S. Gray, J. A. Richards, J. Gilkerson, D. Xu, U. Yapanel, and S. F. Warren, "Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development," *Proceedings of the National Academy of Sciences*, vol. 107, no. 30, pp. 13 354–13 359, 2010.

[26] J. F. Santos, N. Brosh, T. H. Falk, L. Zwaigenbaum, S. E. Bryson, W. Roberts, I. M. Smith, P. Szatmari, and J. A. Brian, "Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 7567–7571.

[27] Y. Kakihara, T. Takiguchi, Y. Ariki, Y. Nakai, S. Takada, Y. Kakihara *et al.*, "Investigation of classification using pitch features for children with autism spectrum disorders and typically developing children," *Am. J. Sign. Process*, vol. 5, pp. 1–5, 2015.