Acoustic Features Characterization of Autism Speech for Automated Detection and Classification

Abhijit Mohanta[1], *Prerana Mukherjee*[2], *Vinay Kumar Mittal*[3]

[1,2]Indian Institute of Information Technology Sri City, Chittoor, Andhra Pradesh, India
[3]*K L University Vijayawada, Andhra Pradesh, India*

Twenty Sixth
National Conference on Communications (NCC)
IIT Kharagpur, 21-23 February 2020

Outline

- Introduction
  - Proposed Plan
  - Significance of this study
- Datasets Collection and Preprocessing
  - Datasets Collection
  - preprocessing
- Speech Production Features
- Classifier's Design
- Results
  - Results using statistical analyses
  - Classification results
  - Comparison with previous studies
- Conclusion

Introduction

- The aim of this study is to differentiate children with *autism spectrum disorder* (ASD) from *normal children*, in terms of their speech production features.

- ASD is a neurodevelopmental disorder which involves communication deficits, social interaction impairments, and hyperfocus or reduced behavioral flexibility [1][*].

- The verbal ASD children often shows some notable acoustic patterns.

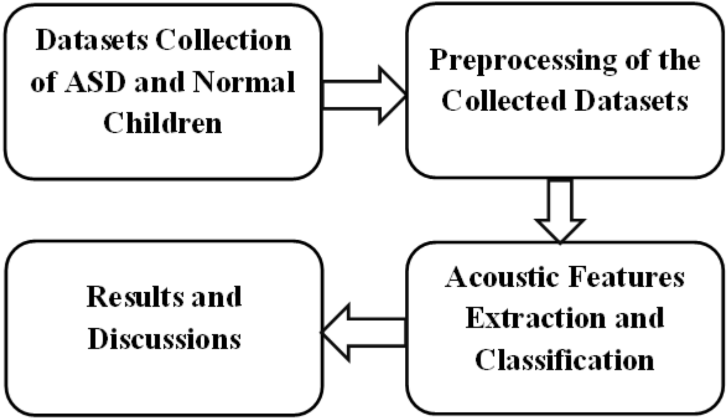
- 1 in 110 children is diagnosed with ASD [2][†].

[*] 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

[†] 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

Introduction (cont.)

Proposed Plan



Datasets Collection of ASD and Normal Children

Preprocessing of the Collected Datasets

Acoustic Features Extraction and Classification

Results and Discussions

Figure 1: Block diagram to represent the basic outline of the proposed plan.

Introduction (cont.)

Why this study is important?

- In our collected datasets all the speakers are non-native (Indian accent) English speakers. Whereas, in the earlier studies like [3], [4], etc., authors have only considered the native English speakers.
- In previous studies datasets were mostly collected from social interactions [2].
- Many robust speech features, especially dominant frequencies (FD1, FD2) [5], strength of excitation (SoE) [6, 7], etc., have not been explored in previous studies.
- Results of this study can be utilized as acoustic markers for ASD diagnosis.

Outline

- Introduction
  - Proposed Plan
  - Significance of this study
- Datasets Collection and Preprocessing
  - Datasets Collection
  - preprocessing
- Speech Production Features
- Classifier's Design
- Results
  - Results using statistical analyses
  - Classification results
  - Comparison with previous studies
- Conclusion

Datasets Collection and Preprocessing

Datasets Collection

Table 1: Datasets details of the (b1) ASD and (b2) Normal children, where (a) represents several attributes and (b) represents statistical measurements

(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

Datasets Collection and Preprocessing (cont.)

preprocessing

1. Signal noise removal:
   - spectral subtraction (SS) [8][‡]
   - minimum mean square error (MMSE) [8]
   - Log MMSE with voice activity detection (VAD) [8]
2. Quantitative measurements:
   - perceptual evaluation of speech quality (PESQ) [9][§]
   - segmental SNR (SNRseg) [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

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

Datasets Collection and Preprocessing (cont.)

Table 2: Quantitative measurements (QM) of several noise cancellation algorithms used in the [A] ASD and the [B] Normal children's speech signal datasets. The SS represents spectral subtraction, MMSE represents minimum mean squire error, and LMV represents Log MMSE_VAD

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

Datasets Collection and Preprocessing (cont.)

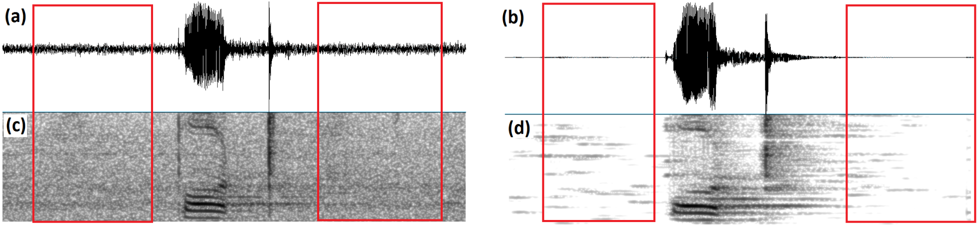


Figure 2: Waveforms and spectrograms of the same speech sound (/dog/) before and after removing noise. The differences in the same speech sound before and after removing its noise can be visualized from the red colored highlighted parts of (c) and (d), and their respective waveforms (a) and (b). The Y-axis in (c) and (d) varies from 0 Hz to 5000 Hz.

Outline

Introduction

- Proposed Plan
- Significance of this study

Datasets Collection and Preprocessing

- Datasets Collection
- preprocessing

Speech Production Features

Classifier's Design

Results

- Results using statistical analyses
- Classification results
- Comparison with previous studies

Conclusion

Speech Production Features

Types of production features used for the classification are:

1. Source Features
   - fundamental frequency (F0) [6]
   - strength of excitation (SoE) [6, 7]
2. Vocal tract filter features
   - dominant frequencies (FD1, FD2) [5]
   - first five formants (F1 to F5) [10]
3. Source-system combined features
   - signal energy (E) [11]
   - zero-crossing rate (ZCR) [12]
   - mel-frequency cepstral coefficients (MFCC) [13]
   - linear prediction cepstrum coefficients (LPCC) [13]

Outline

Introduction

- Proposed Plan
- Significance of this study

Datasets Collection and Preprocessing

- Datasets Collection
- preprocessing

Speech Production Features

Classifier's Design

Results

- Results using statistical analyses
- Classification results
- Comparison with previous studies

Conclusion

Classifier's Design

Six different classifiers are utilized in this study:

- support vector machine (SVM) [14]
- k-nearest neighbors (KNN) [15]
- linear discriminant (LD) [16]
- quadratic discriminant (QD) [17]
- decision tree (DT) [18]
- logistic regression (LR) [19]

Outline

- Introduction
  - Proposed Plan
  - Significance of this study
- Datasets Collection and Preprocessing
  - Datasets Collection
  - preprocessing
- Speech Production Features
- Classifier's Design
- **Results**
  - Results using statistical analyses
  - Classification results
  - Comparison with previous studies
- Conclusion

Results

Results using statistical analyses

Table 3: The [A] mean and [B] SD values of the acoustic (a) features; (b) and (d) represent the values of acoustic features of the *ASD* children, and (c) and (e) represent the values of acoustic features of the *Normal* children. Values of frequencies F0, F1-F5 and FD1-FD2 are in Hz

(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

Results (cont.)

Results of statistical analyses

A few key observations are (*Details results are given in the paper*):

- The ASD children have higher μ_{F0}, μ_{E}, and μ_{ZCR} values than the normal children.
- The ASD children have lower μ_{SoE} value than the normal children.
- VT filter features μ_{F2}, μ_{F3}, μ_{F4}, and μ_{F5}, have higher values for ASD children than the normal children.
- But μ_{F1} has lower value for ASD children than the normal children.
- The μ_{FD1} have higher value and μ_{FD2} have lower value for the ASD children as compared with the normal children.

Results (cont.)

Classification results

Table 4: 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

Results (cont.)

Comparison with previous studies

Table 5: Comparison of our results with similar previous studies: (a) represents authors, (b) represents classifier's name, and (c) represents classification accuracy in percentage

(a) Authors	(b) Classifiers	(c) Accuracy (%)
Fusaroli et al., [20]	QD, linear regression	86.0
Oller et al., [3]	LD analysis	86.0
Santos et al., [2]	SVM	79.1
Kakihara et al., [21]	SVM	74.9
Santos et al., [2]	probabilistic neural network (PNN)	97.7
Proposed method	*KNN*	*96.5*

Outline

Introduction

- Proposed Plan
- Significance of this study

Datasets Collection and Preprocessing

- Datasets Collection
- preprocessing

Speech Production Features

Classifier's Design

Results

- Results using statistical analyses
- Classification results
- Comparison with previous studies

Conclusion

Conclusion

- It is observed that there are significant differences between the Indian ASD and the normal children, in terms of their speech production characteristics.
- The results obtained in this work can be utilized as an acoustic biomarker to identify ASD from the speech signal at a very early age.
- These robust results obtained from Indo English children with ASD can be compared with native English children with ASD, in future studies.
- A small size of speech data especially for female ASD children is a limitation of this research work.

References I

[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] 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.

[3] 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.

[4] 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.

[5] V. K. Mittal and B. Yegnanarayana, "Study of characteristics of aperiodicity in noh voices," *The Journal of the Acoustical Society of America*, vol. 137, no. 6, pp. 3411–3421, 2015.

[6] 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.

[7] B. Yegnanarayana and K. S. R. Murty, "Event-based instantaneous fundamental frequency estimation from speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 614–624, 2009.

[8] 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.

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

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

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

References II

[12] 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.

[13] 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.

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

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

[16] 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.

[17] 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.

[18] 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.

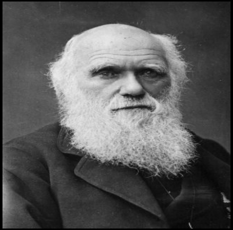
[19] R. Greiner, X. Su, B. Shen, and W. Zhou, "Structural extension to logistic regression: Discriminative parameter learning of belief net classifiers," *Machine Learning*, vol. 59, no. 3, pp. 297–322, 2005.

[20] 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.

[21] 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.

Thank You.

A famous person with autism!



Charles Darwin