Prosody Features Characterization of Autism Speech for Automated Detection and Classification

Abhijit Mohanta
MS by Research Student
Roll No: 201681001

Indian Institute of Information Technology, Sri City

10-Dec-2019

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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
- Classification results
- Comparison with previous studies

Conclusion

1 Introduction
- Proposed Plan
- Significance of this study

2 Datasets Collection and preprocessing
- Datasets Collection
- preprocessing

3 Speech Production Features

4 Classifier's Design

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

6 Conclusion

7 References

Introduction

- The aim of this research is to differentiate children with *Autism spectrum disorder* (ASD) from *normal children*, in terms of their speech production features.
- ASD is a set of neurodevelopmental disorders that are defined by social communication impairments and reciprocity [1]*.
- 1 in 68 children affected with autism was reported in 2014 [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

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Introduction (cont.)

Proposed Plan

This study is divided into four major steps.

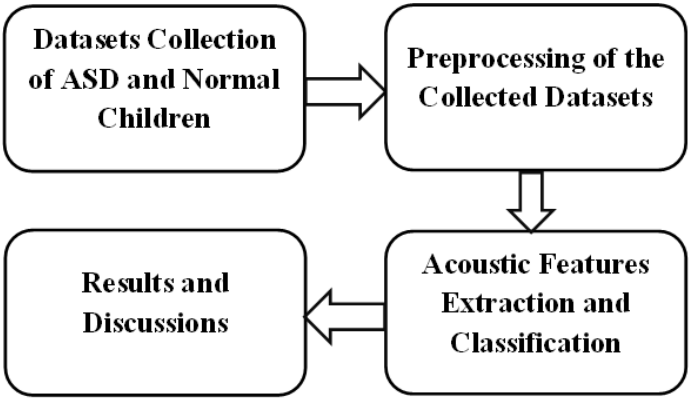


Figure 1: Proposed plan's block diagram.

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Introduction (cont.)

Why this study is important?

- In the collected datasets all the speakers are *non-native (Indian accent) English speakers*. But, 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 interaction [2], constrained production [5], etc.
- Many robust speech features, especially dominant frequencies (FD1, FD2) [6], strength of excitation (SoE) [7] and fifth formant frequency (F5) [8] have not been explored in previous studies.
- Results of this study can be used as acoustic markers for ASD.

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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
- Classification results
- Comparison with previous studies

Conclusion

1 Introduction
- Proposed Plan
- Significance of this study

2 Datasets Collection and preprocessing
- Datasets Collection
- preprocessing

3 Speech Production Features

4 Classifier's Design

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

6 Conclusion

7 References

Datasets Collection and preprocessing

Datasets Collection

(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

Table 1: 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

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Datasets Collection and preprocessing I

preprocessing

1. Signal noise removal:
   - spectral subtraction (SS) [9][‡]
   - minimum mean square error (MMSE) [9]
   - Log MMSE with voice activity detection (VAD) [9]
2. Quantitative measurements:
   - perceptual evaluation of speech quality (PESQ) [10][§]
   - segmental SNR (SNRseg) [10]

[‡] 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

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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

Table 2: 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

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

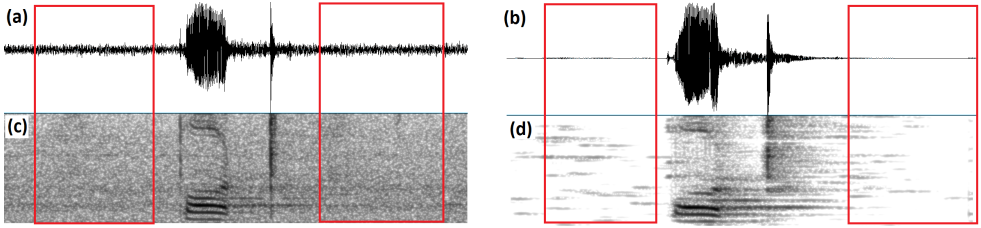


Figure 2: Waveform and spectrogram of a same word (/tom/) before and after removing the signal noise. Here, (a) and (b) represent two waveforms before and after removing the signal noise, respectively, and (c) and (d) represent spectrograms of those two signals (in (a) and (b)) before and after removing the signal noise, respectively.

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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
- Classification results
- Comparison with previous studies

Conclusion

1 Introduction
- Proposed Plan
- Significance of this study

2 Datasets Collection and preprocessing
- Datasets Collection
- preprocessing

3 Speech Production Features

4 Classifier's Design

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

6 Conclusion

7 References

Speech Production Features

Types of production features used for the classification are:

1. Source Features
   - fundamental frequency i.e., F0 [7]
   - strength of excitation (SoE) [7]
2. Vocal tract system features
   - dominant frequencies (FD1, FD2) [6]
   - first five formants (F1 to F5) [11]
3. Source-system combined features
   - signal energy (E) [12]
   - zero-crossing rate (ZCR) [13]
   - mel-frequency cepstral coefficients (MFCC) [14]
   - linear prediction cepstrum coefficients (LPCC) [14]
   - VGGish audio features [15]

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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
- Classification results
- Comparison with previous studies

Conclusion

1 Introduction
- Proposed Plan
- Significance of this study

2 Datasets Collection and preprocessing
- Datasets Collection
- preprocessing

3 Speech Production Features

4 Classifier's Design

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

6 Conclusion

7 References

Classifier's Design

Six different classifiers are utilized in this study:

- support vector machine (SVM) [16]
- K-nearest neighbors (KNN) [17]
- linear discriminant (LD) [18]
- quadratic discriminant (QD) [19]
- decision tree (DT) [20]
- logistic regression (LR) [21]

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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
- Classification results
- Comparison with previous studies

Conclusion

1 Introduction
- Proposed Plan
- Significance of this study

2 Datasets Collection and preprocessing
- Datasets Collection
- preprocessing

3 Speech Production Features

4 Classifier's Design

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

6 Conclusion

7 References

Results

Results using statistical analyses

(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

Table 3: 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

endcenter

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Results (cont.)

Results using statistical analyses

Some of the key observations are:

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

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Results (cont.)

Classification results

(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

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

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Results (cont.)

Classification results

(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

Table 5: *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

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Results (cont.)

Comparison with previous studies

(a) Authors	(b) Classifiers	(c) Accuracy (%)
Fusaroli et al., [22]	QD, linear regression	86.0
Oller et al., [3]	LD analysis	86.0
Santos et al., [2]	SVM	79.1
Kakihara et al., [23]	SVM	74.9
Santos et al., [2]	probabilistic neural network (PNN)	97.7
Proposed method	*KNN*	*96.5*

Table 6: Comparison of our results with some of the previous studies.

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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
- Classification results
- Comparison with previous studies

Conclusion

1 Introduction
- Proposed Plan
- Significance of this study

2 Datasets Collection and preprocessing
- Datasets Collection
- preprocessing

3 Speech Production Features

4 Classifier's Design

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

6 Conclusion

7 References

Conclusion

- It is observed that there are significant differences between the 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 for female ASD children is a limitation of this research work.

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

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] D. Bone, M. P. Black, C.-C. Lee, M. E. Williams, P. Levitt, S. Lee, and S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[6] 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.

[7] 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.

[8] 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.

[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.

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

References II

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

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

[13] 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.

[14] 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.

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

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

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

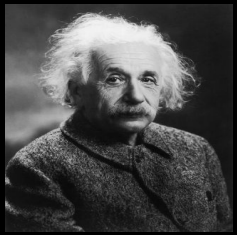
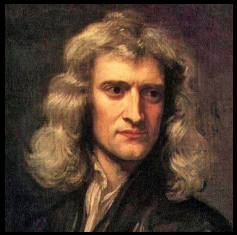
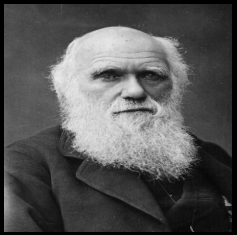
[22] R. Fusaroli, D. Bang, and E. Weed, "Non-linear analyses of speech and prosody in asperger's syndrome," in *International Meeting For Autism Research*, 2013.

[23] 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.

1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References

Thank You...

Some famous People with autism!


1 Introduction
Proposed Plan
Significance of this study
2 Datasets Collection and preprocessing
Datasets Collection
preprocessing
3 Speech Production Features
4 Classifier's Design
5 Results
Results using statistical analyses
Classification results
Classification results
Comparison with previous studies
6 Conclusion
7 References