

Machine Learning Classification of Natural Conversational Utterances Using Acoustic Features Drawn from Children with ASD and Typical Controls

Penn Medici Medicine

Sunghye Cho¹, Mark Liberman¹, Neville Ryant¹, Keith Bartley², Robert T. Schultz^{2,3}, & Julia Parish-Morris^{2,3} ¹ University of Pennsylvania, Linguistic Data Consortium, ² Children's Hospital of Philadelphia Center for Autism Research, ³ University of Pennsylvania Perelman School of Medicine

Background

- Earliest descriptions of ASD: atypical speech patterns, including unusual prosody [1], [2]
- Previous research on phonetic properties of speech in ASD: mostly elicited in a highly structured context with an autism expert. While valuable, results may not generalize to the everyday conversations that really matter for children with ASD.
- Goal: Develop a machine learning classifier approach to children's natural interactions with a naïve conversational partner.

Feature extra

9 pitch features, F0 mean, IQR, M

16 intensity featur as mean intensity

8 voice-quality fe such as CPP, H1-ł

4 duration features word duration, spe

Data

- 35 ASD (12 boys, mean age = 11.42 yrs) and 35 typical controls (TC; 14 boys, mean age = 10.57 yrs) who are matched on age, sex ratio, and full-scale IQ values
- 5-minute "get-to-know-you" conversation with a naïve conversational partner who was not aware of participants' diagnostic status
- Conversations were audio/video recorded and annotated & time-aligned by a team of trained annotators.
- Turns with overlapping speech were excluded.

Turn-level classification results (mean of all CVs; percent values):

Accuracy	Precision	Recall	F1-score
60%	60%	58%	60%

Accuracy	Precision	Recall	F1-score
70%	72%	66%	69%

Contact Sunghye Cho (csunghye@sas.upenn.edu) for questions!

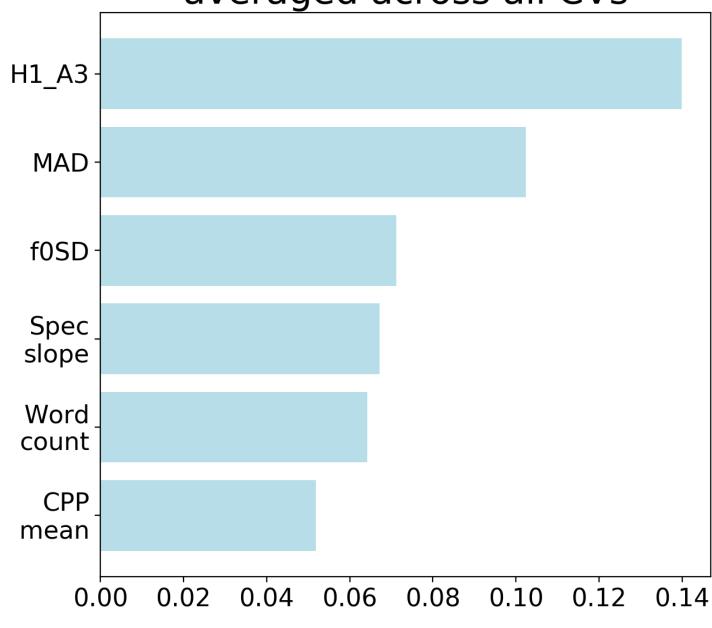
Pipeline

action	Feature processing	Turn-level
including IAD, jitter	Pitch-tracking twice to reduce pitch-tracking errors	Gradient Boosting Model + leave-one-group-out CV
res, such and HNR	Simple imputation with median values	
octurac	Eastura coloction using	
eatures, H2, H1-A3	Feature selection using correlation (> 0.6)	Speaker-level
s, such as eech rate	Standardization with StandardScaler	Majority vote with the turn- level model's prediction

Classification results

• Speaker-level classification results:

• 23 correct prediction out of 35 ASD (= 66%) • 26 correct prediction out of 35 TC (= 74%)



Feature importance averaged across all CVs

Discussion & future direction

Children's Hospital

Center for Autism Research

of Philadelphia

- Our classifier shows that acoustic features lacksquarefrom brief natural conversations are useful for distinguishing children with ASD and TC.
- The classifier finds that voice-quality and pitch-related features are most important in identifying children with ASD and TC.
- The result is promising given that current data are drawn from natural conversations, which tend to be messier and more variable than other types of data.
- Future direction:
 - Use a more sophisticated feature selection methods
 - Include lexical information, for example word choice, filled pauses
 - Collect more data

Acknowledgement

- We thank the children and families that participated in our research, as well as clinicians and staff at the Center for Autism Research.
- This study was supported by an Autism Science Foundation Postdoctoral Fellowship to JPM; the Eagles Charitable Trust, McMorris Family Foundation, and Allerton Foundation to RTS; and NICHD 5U54HD086984-03 to Michael B Robinson & RTS.

References

- [1] H. Asperger, "Die Autistischen Psychopathen im Kindesalter," Arch. Psych. Nervenkrankh, vol. 117, pp. 76-136, 1944.
- [2] L. Kanner, "Autistic disturbances of affective contact." Nerv. Child, vol.2, pp. 217-250, 1943.