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

Sunghye Cho1, Mark Liberman1, Neville Ryant1, Keith Bartley2, Robert T. Schultz2,3, & Julia Parish-Morris2,3

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.

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.

Classification results

- Turn-level classification results (mean of all CVs; percent values):
  
<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>60%</td>
<td>60%</td>
<td>58%</td>
<td>60%</td>
</tr>
</tbody>
</table>

- Speaker-level classification results:
  
<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>70%</td>
<td>72%</td>
<td>66%</td>
<td>69%</td>
</tr>
</tbody>
</table>

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

Discussion & future direction

- Our classifier shows that acoustic features from 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 to Michael B Robinson & RTS.

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

References
