

# PERCEPT: A database of clinical child speech for automatic speech recognition and classification

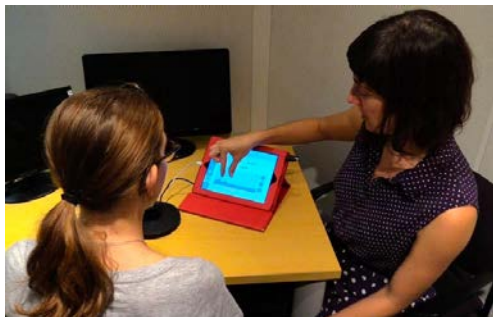
Tara McAllister<sup>1</sup>, Nina Benway<sup>2</sup>, Asif Salekin<sup>2</sup>, Jonathan L.  
Preston<sup>2</sup>, Elaine Hitchcock<sup>3</sup>

<sup>1</sup>*New York University*, <sup>2</sup>*Syracuse University*, <sup>3</sup>*Montclair State University*



## Childhood speech sound disorder

- ▶ Children with speech sound disorder (SSD) exhibit speech deviations that make them less intelligible than their peers.
- ▶ Creates a barrier to academic and social participation, with potentially life-long educational/occupational ramifications.
- ▶ Intervention can reduce these impacts, but poor outcomes are common due to insufficient dosage of treatment or lack of access to effective tools.



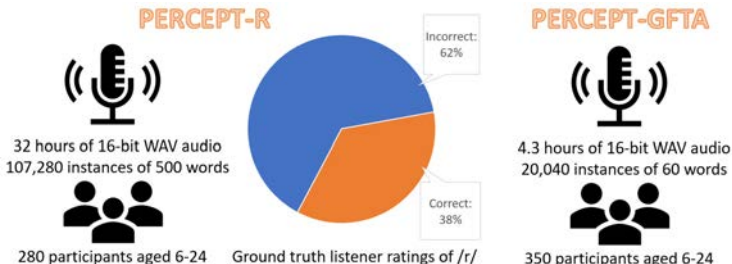
## AI-enhanced intervention

- ▶ Treatment of speech sound disorder could be enhanced through the development of AI/ML tools.
  - ▶ For example, applications with automated scoring could be used to augment clinician services.
  - ▶ Higher-intensity practice → greater progress in therapy.
- ▶ However, no computerized treatment to date has demonstrated sufficient accuracy for clinical use with children.
- ▶ Fundamental barrier: Lack of large-scale speech corpora from child and clinical speakers.



## Corpus creation

- ▶ Drawing on audio from 10+ years of previous research, we curated two corpora:
  - ▶ PERCEPT-R: Focused on children/adolescents with SSD affecting American English /r/.
  - ▶ PERCEPT-GFTA: Children/adolescents with diverse speech errors producing a standardized articulation measure.
- ▶ All records are linked to ground-truth orthographic labels. PERCEPT-R records are labeled with blinded listener judgments of /r/ accuracy.



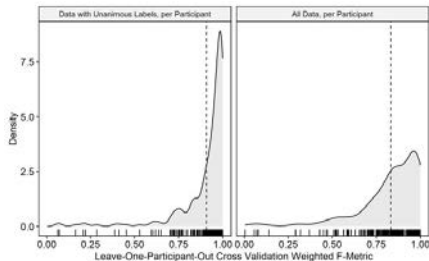
# Corpus dissemination

- ▶ Partnered with [PhonBank](#) (Rose & MacWhinney, 2014), an NIH-funded open-access data-sharing platform for speech and language research.
  - ▶ [PERCEPT-GFTA](#)
  - ▶ [PERCEPT-R](#)
- ▶ Audio records are time-aligned with searchable phonetic transcriptions through Phon software.
- ▶ Also releasing an AI-ready version (Python-compatible data structure) under [derived corpora](#).



# Application

- ▶ PERCEPT-R has been used to train a classifier ([Benway et al., 2022](#)) that automatically labels children's /r/ sounds as correct or incorrect.
- ▶ Testing has yielded F1-scores above .8 and processing times below 2 seconds, and clinical testing is underway.
- ▶ Corpora can be used to fine-tune pretrained models for child/clinical speech applications.



Thank you!

[tkm214@nyu.edu](mailto:tkm214@nyu.edu)  
[@BITSLabNYU](#)

Parent grant: NIH R01DC017476 (PI: McAllister).

