NRI:INT:COLLAB Development, Deployment and Evaluation of Personalized Learning Companion Robots for Early Literacy and Language Learning

Cynthia Breazeal (PI), Hae Won Park (co-PI) *Abeer Alwan (co-PI), Alison Bailey* *Mari Ostendorf*

MIT UCLA W

Award Number: NSF 1734443/1734380

Award Date: 9/2017-8/2021

The Challenge

P1: 1/3 of American children do not reaching basic levels of literacy, and 2/3 fail to reach proficiency levels of literacy.

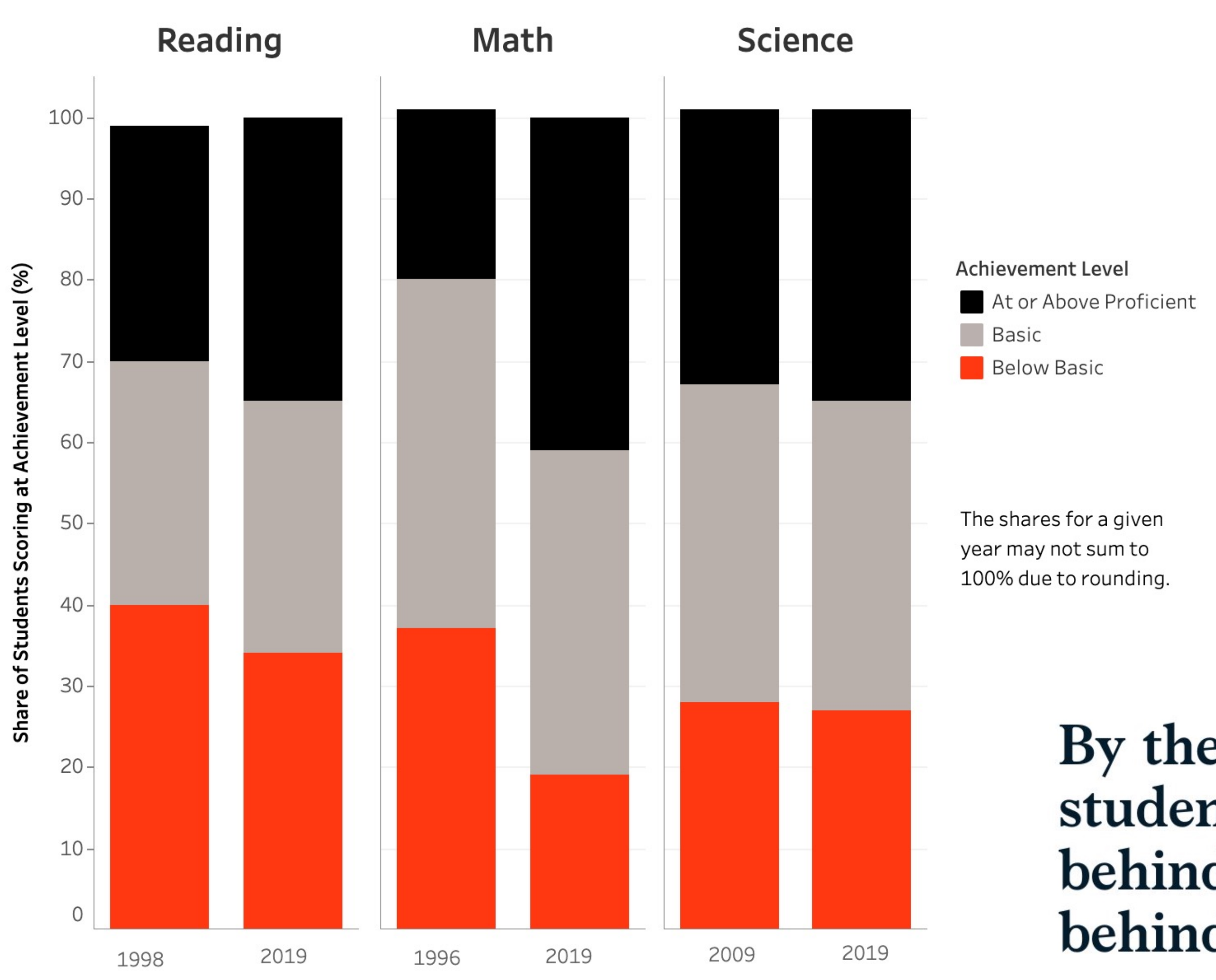
→ PreK/Kindergarten is the most critical and cost-effective time to intervene.

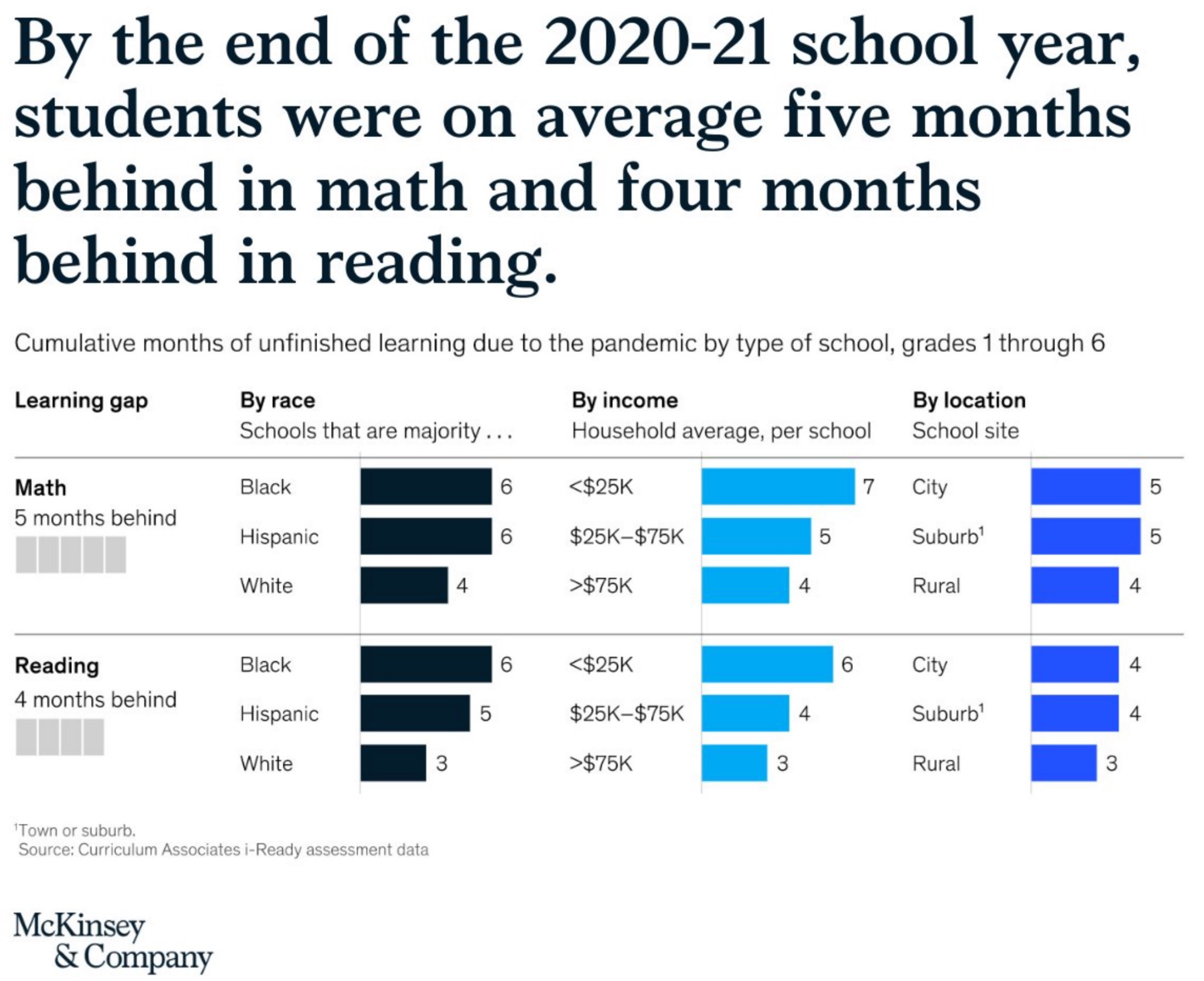
→ Learning outcomes has been worsened by COVID-19 pandemic

P2: Many intelligent tutoring systems have been proposed, but neglect engaging young children's social and emotional learning abilities and only focus on inserting knowledge.

→ Personalized, social robot augmented learning interventions that are well matched to the social, emotional, and cognitive learning needs of young children could dramatically improve school readiness.

I-02a: Fourth Graders' Performance on the National Assessment of Educational Progress (NAEP), by Subject, 1996/1998/2009–2019*





Scientific Impact Goals

Impact #1: Long-term Personalized Reading Companion

- Jibo Stations sent to **children's homes** to **support remote learning** during COVID pandemic.
- **Cross-task Learning** to accelerate personalization of multiple literacy tasks.
- **Affective Personalization** for maximizing engagement and learning using hierarchical reinforcement learning.

16 in classrooms and 12 at homes

Question Proposal

Verbal & Non-verbal

Training Data

Real-time ASR

Training Data

Impact #3: Contextually Grounded Dialogic QnA

- Accounting for child engagement and uncertainty in **question timing**
- Contextually grounded dynamic **question generation** and quality assessment
- **Novel corpora** of disfluency annotated child speech and questions designed for spoken conversations

txt/audio Time Alignment

Impact #2: Automatic Child Speech Recognition

- **Engaging speech collection protocol** administered by a social robot and a **novel longitudinal corpus** of child speech
- **Effective child ASR system** using **transfer learning** & **data augmentation** techniques
- **Diarization** and **speaker identification** systems to enable personalized learning and assessment

Broader Impacts

- Advance computational methods, integrated systems, and human factors insights for child-centered conversational systems
- New AI-enabled assessment and intervention methods to improve learning outcomes for early literacy that is ultimately more scalable and cost-effective
- Personalized learning companions that augment, extend, and support parents and teachers in early literacy goals in school and home settings
- Help to address early literacy learning gaps for young children due to the pandemic

Automatic Reading Skill Assessment & Learning Content Personalization

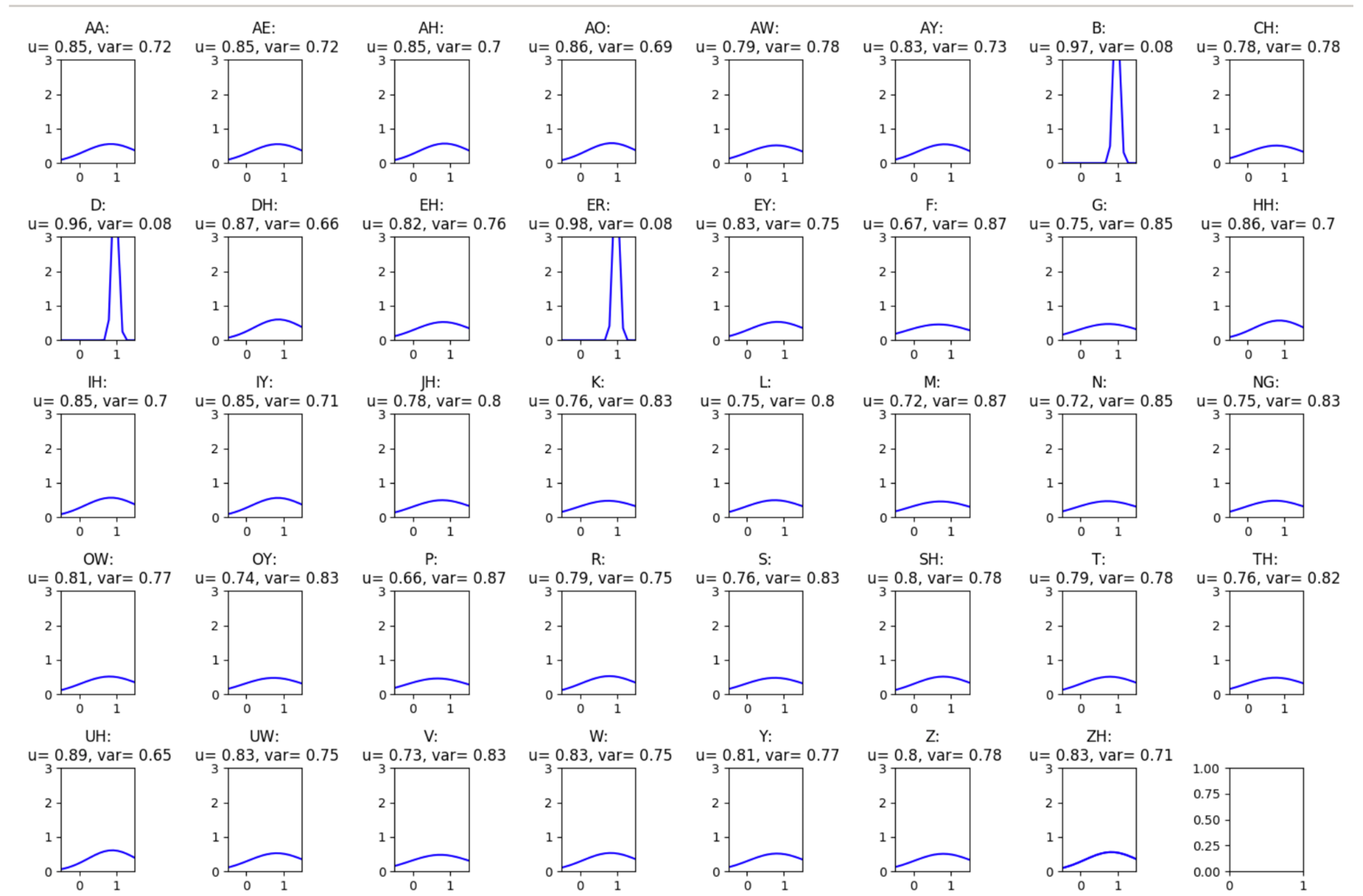
Scientific Impact

Automatic Reading Skill Assessment and Learning Content Personalization

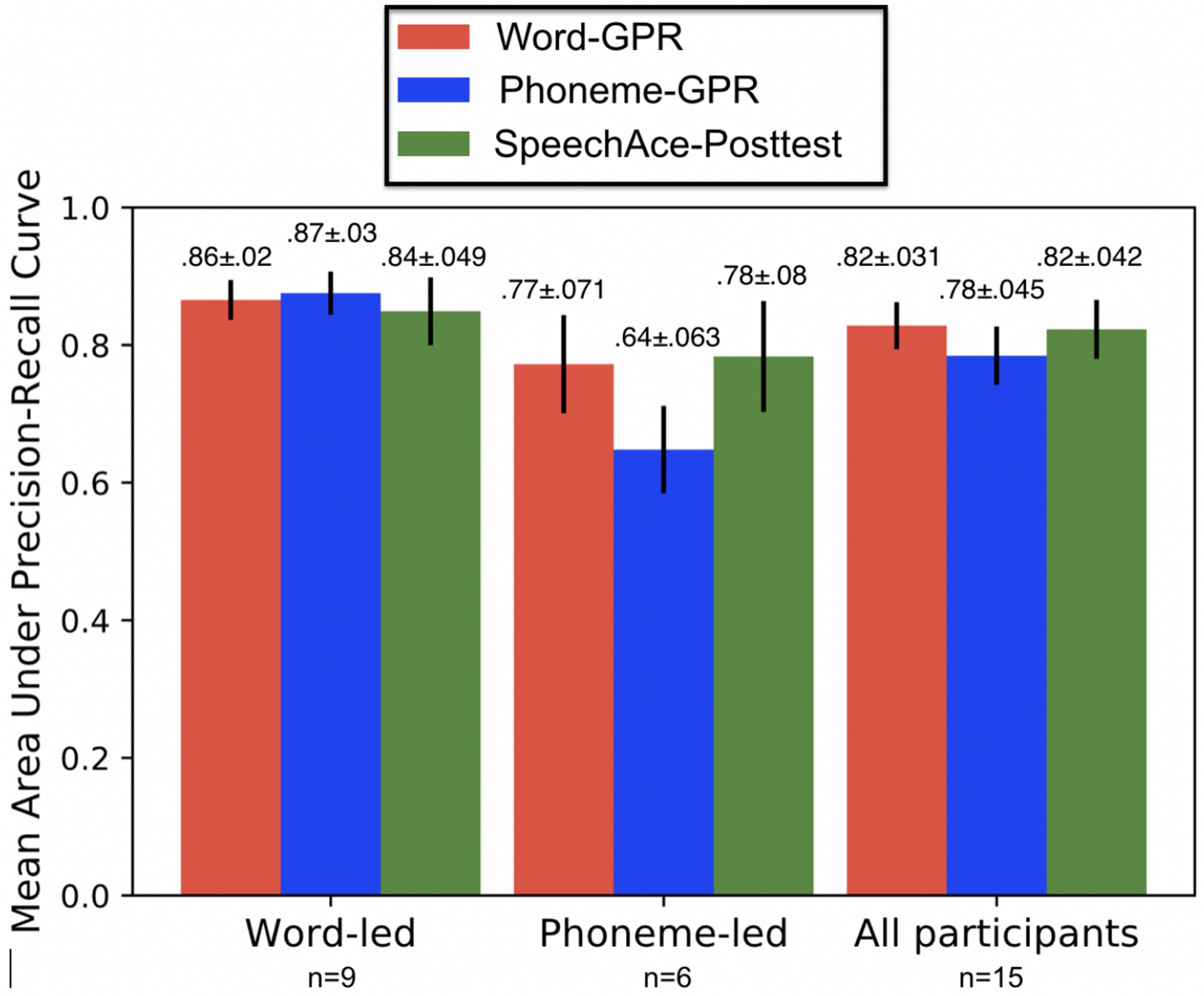
Personalized intervention is known to be the most effective method for early literacy and language learning but it is hard to achieve in classroom setting

Innovation

- **Word- and phoneme-level student models** based on Gaussian Process Regression (GPR) guides personalized learning content selection.
- An **active learning protocol** for efficiently determining and introducing adaptive, personalized content.
- An interactive tablet game (WordRacer) that facilitates child-robot word pronunciation game play



Posterior GPR Phoneme Model



Word and Phoneme-level GPR Performance

S. Spaulding, H. Chen, S. Ali, M. Kulinski, and C. Breazeal, “A social robot system for modeling children’s word pronunciation: Socially interactive agents track,” AAMAS 2018.

I. Grover, H. W. Park, and C. Breazeal, “A semantics-based model for predicting children’s vocabulary,” in IJCAI 2019.

Robot Behavior Policy Personalization for Maximizing Student Engagement

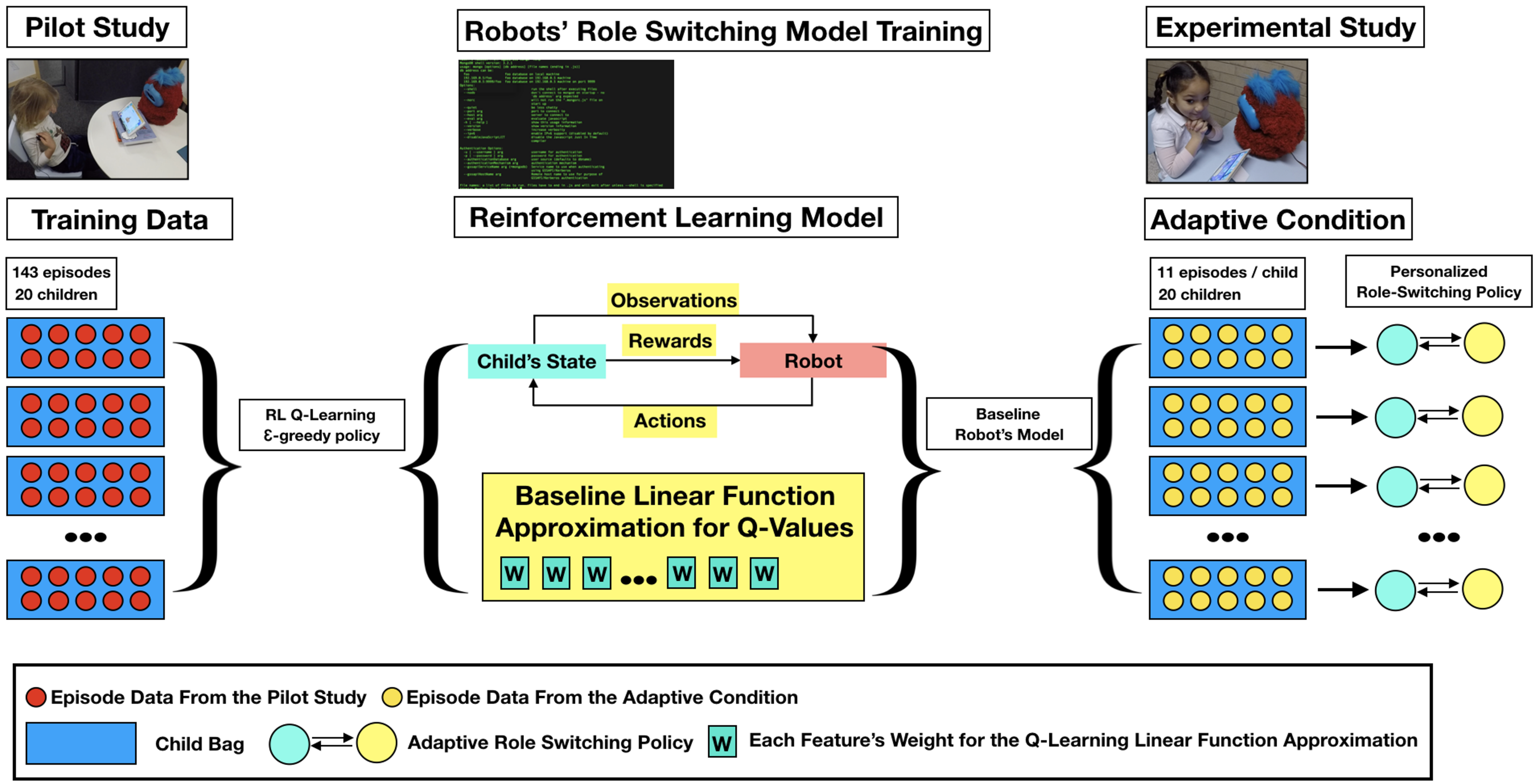
Scientific Impact

Personalizing Robot Behavior to Maximize Engagement

Each student learns and is motivated differently, so the robot learning companion should personalize its behavior policy to maximize each student's engagement.

Innovation

- Robot learns **RL based personalized role-switching behavior policy** that maximizes each child's learning performance
- **Models for the robot's different collaborative roles** (e.g., tutor, tutee, peer) and a set of behaviors associated with each role (e.g., question asking, encouragement, providing information, etc.)
- An interactive tablet game (WordQuest) that facilitates child-robot peer-to-peer word learning



H. Chen, H. W. Park, and C. Breazeal, "Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement," Computers & Education, 2020.

H. Chen, H. W. Park, X. Zhang, and C. Breazeal, "Impact of interaction context on the student affect learning relationship in child-robot interaction," in HRI 2020.

Robot Behavior Policy Personalization for Maximizing Student Engagement

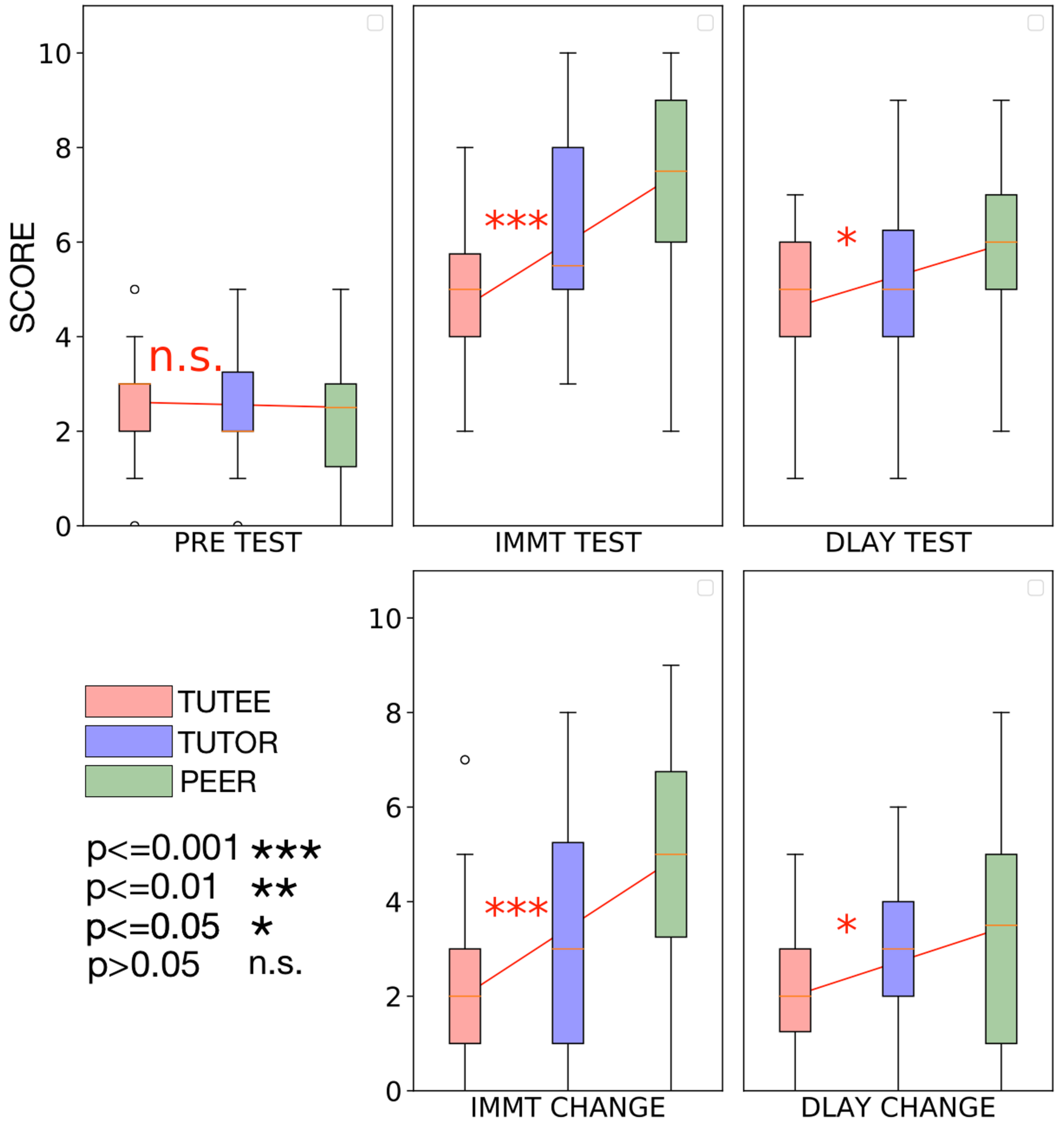
Scientific Impact

Personalizing Robot Behavior to Maximize Engagement

Each student learns and is motivated differently, so the robot learning companion should personalize its behavior policy to maximize each student's engagement.

Innovation

- Robot learns **RL based personalized role-switching behavior policy** that maximizes each child's learning performance
- **Models for the robot's different collaborative roles** (e.g., tutor, tutee, peer) and a set of behaviors associated with each role (e.g., question asking, encouragement, providing information, etc.)
- An interactive tablet game (WordQuest) that facilitates child-robot peer-to-peer word learning



Trend Analysis of word test scores shows the significant advantage of the adaptive role-switching policy (PEER condition).

H. Chen, H. W. Park, and C. Breazeal, "Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement," Computers & Education, 2020.

H. Chen, H. W. Park, X. Zhang, and C. Breazeal, "Impact of interaction context on the student affect learning relationship in child-robot interaction," in HRI 2020.

Context-Aware Affective Personalization for Reinforcement Learning Agents using Reward Shaping

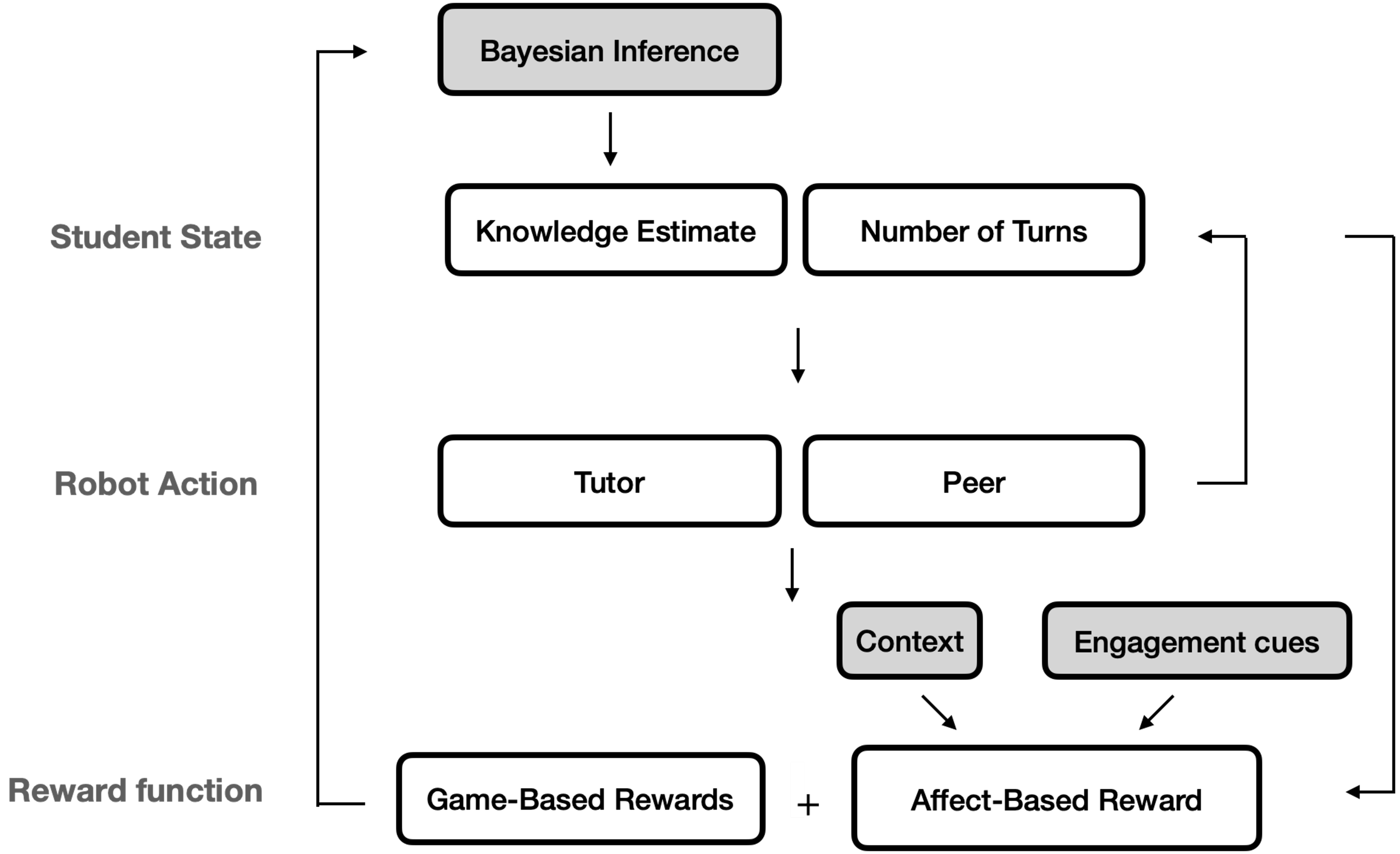
Scientific Impact

Personalizing Robot Behavior to Maximize Learning Outcomes

Students show different facial affective cues when learning new information. These affective cues can be used to guide robot's role switching policy to maximize learning outcomes.

Innovation

- Robot approximates **optimal role-switching policy** while **incorporating facial affective engagement cues** to maximize a student's learning gains.
- Rewards from affective features are **conditioned on the context** in which behaviors are exhibited.
- Student's knowledge is estimated using Bayesian inference and further incorporated into the learning algorithm.



Transferrable Multi-task Personalized Student Models

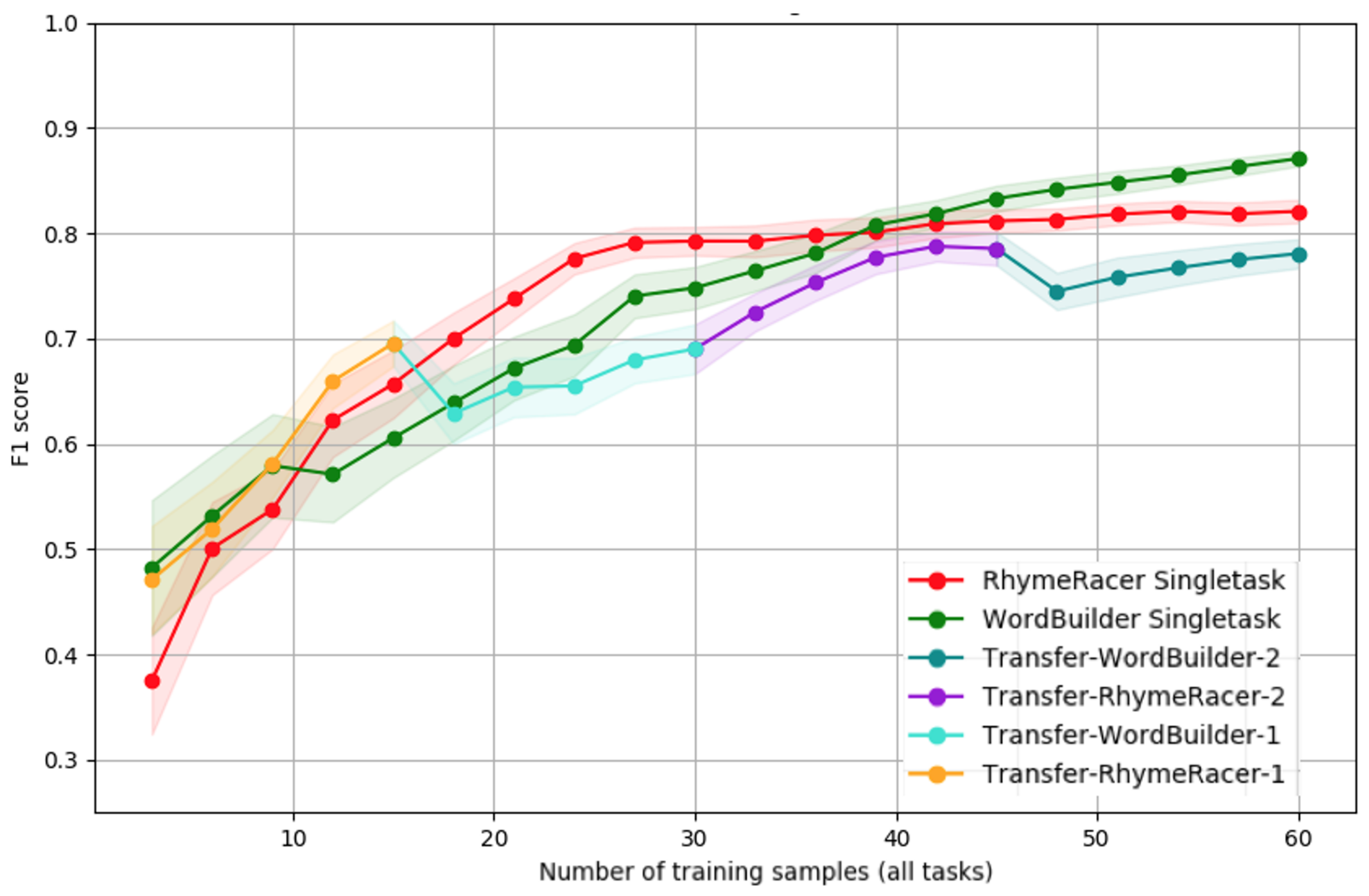
Scientific Impact

Personalized student models have been shown to improve student learning and engagement outcomes

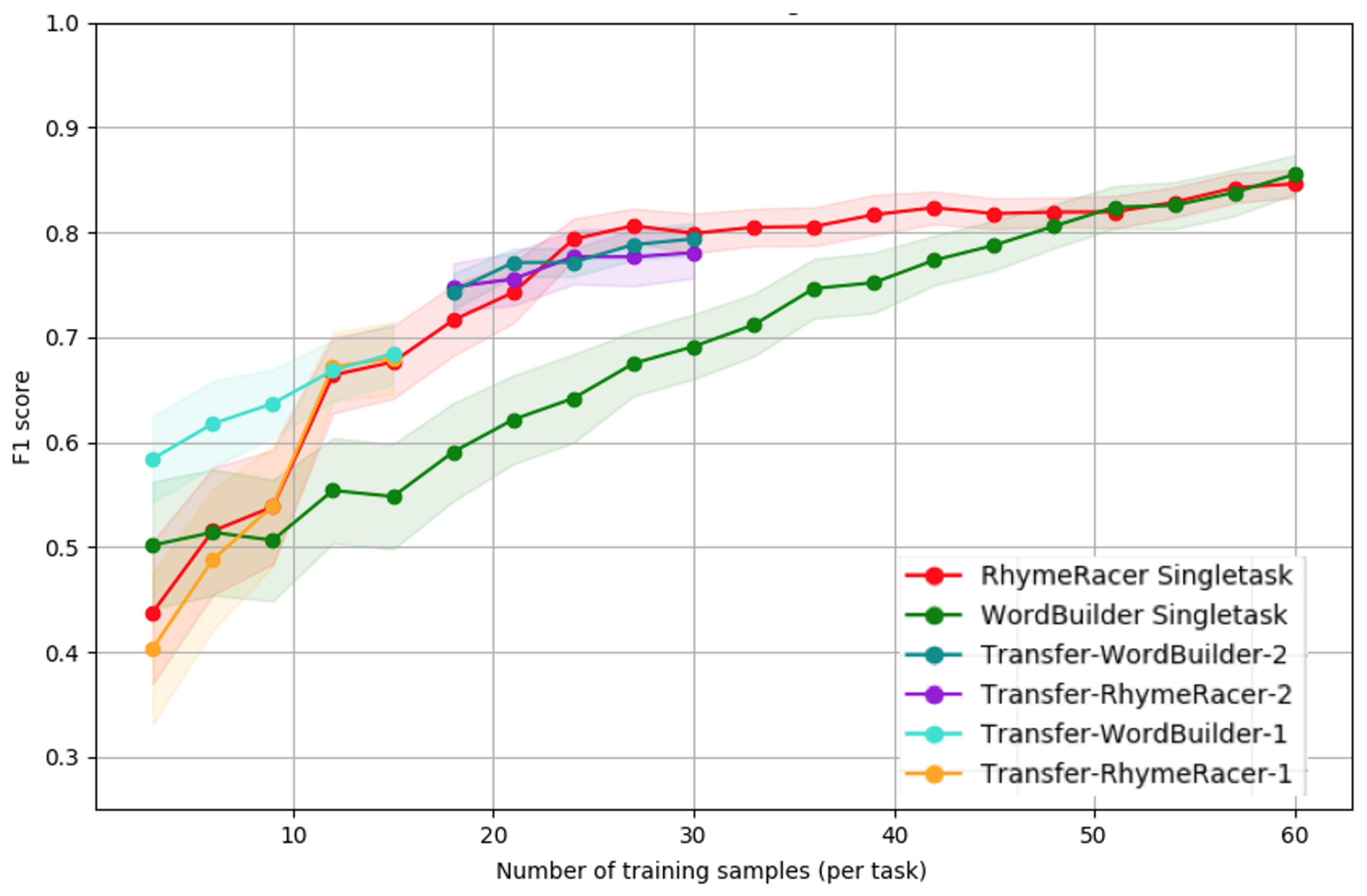
Limited data and narrowly-focused interactions are two challenges researchers face in developing deeply personalized models

Innovation

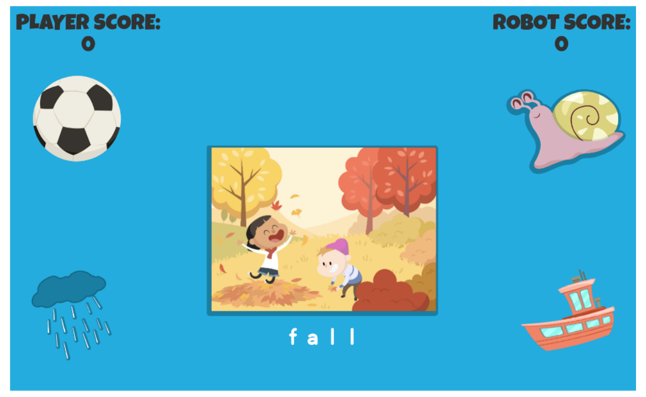
- **“Multitask Personalization”**
  a paradigm for designing student models that are **transferrable** across tasks, improving model data efficiency and domain flexibility.
- **Gaussian Process-based model** set in a joint word-space domain gives two game tasks a shared representation
- **Instance Weighting protocol** transfers prior data based on task-similarity w.r.t observed data points



Transfer Model trades off small reduction in avg performance for dual-task applications with same amount of cumulative data

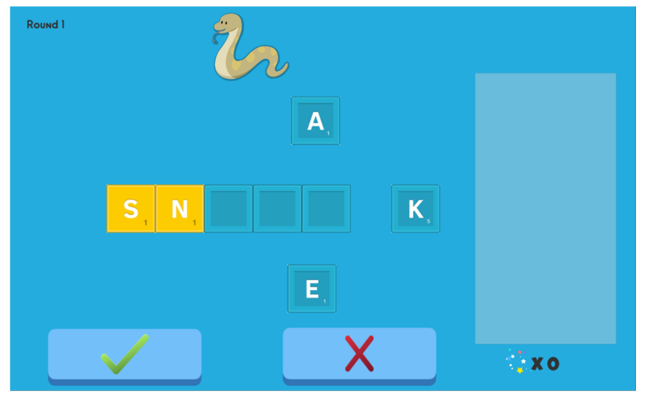


Transfer Model strongly outperforms single-task model with same amount of target-task data



Task 1
phonetic awareness

Task 2
word spelling



Spaulding, S., Shen, J., Park, H., and Breazeal, C., “Towards transferrable personalized student models in educational games,” AAMAS 2021.

Spaulding, S., Shen, J., Park, H. W., & Breazeal, C. “Lifelong Personalization via Gaussian Process Modeling for Long-Term HRI,” *Frontiers in Robotics and AI*, *8*, 152. 2021.

Automatic Recognition of Child Speech

Abeer Alwan

Electrical and Computer Engineering Department, UCLA

Why is Child Speech Recognition Important?

- The use of interactive technology is rapidly increasing
  - Young children rely on speech to interact with computers

- Child speech recognition is used for several applications including:
  - Assessment in classroom environments
  - Educational games
  - Clinical diagnosis

Difficulties of Child ASR

- Lack of publicly available child speech data.
  - Adult databases are typically used for training child ASR
  - There is major acoustic mismatch between adult and child speech, hence normalization is often required.

- Child speech exhibits larger inter- and intra-speaker variability (compared to adults).

- **Kennedy et al. (2018):** 15% WER on child (4-6 years) digit recognition using state-of-the-art ASR APIs (Google, Nuance, Bing, Sphinx).
  - In contrast, adult digit recognition WER is < 2%

Progress at UCLA

1. Collected 55 hours of Child Speech
2. Developed ASR algorithms via:

   Novel Frequency Normalization techniques

   Novel Data Augmentation Techniques

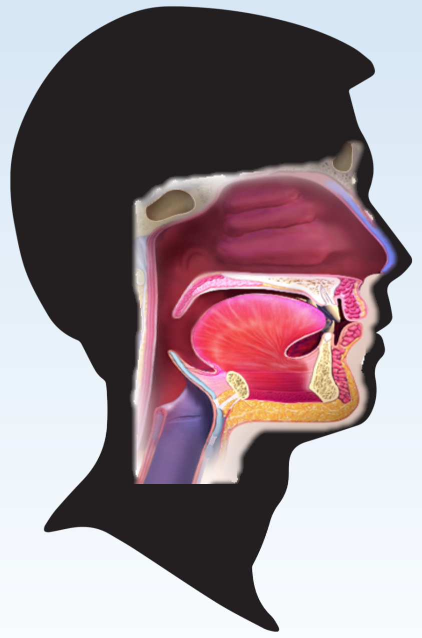
   Novel Self-Supervised Learning Techniques

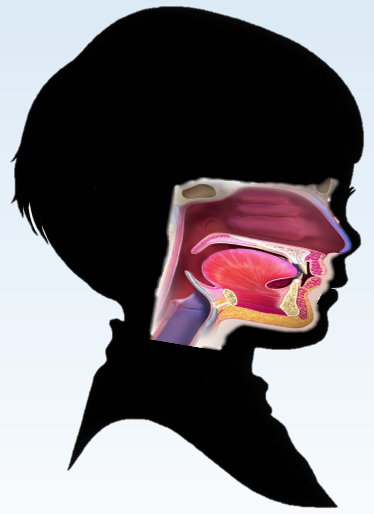
1. The JIBO Kids Database

Database

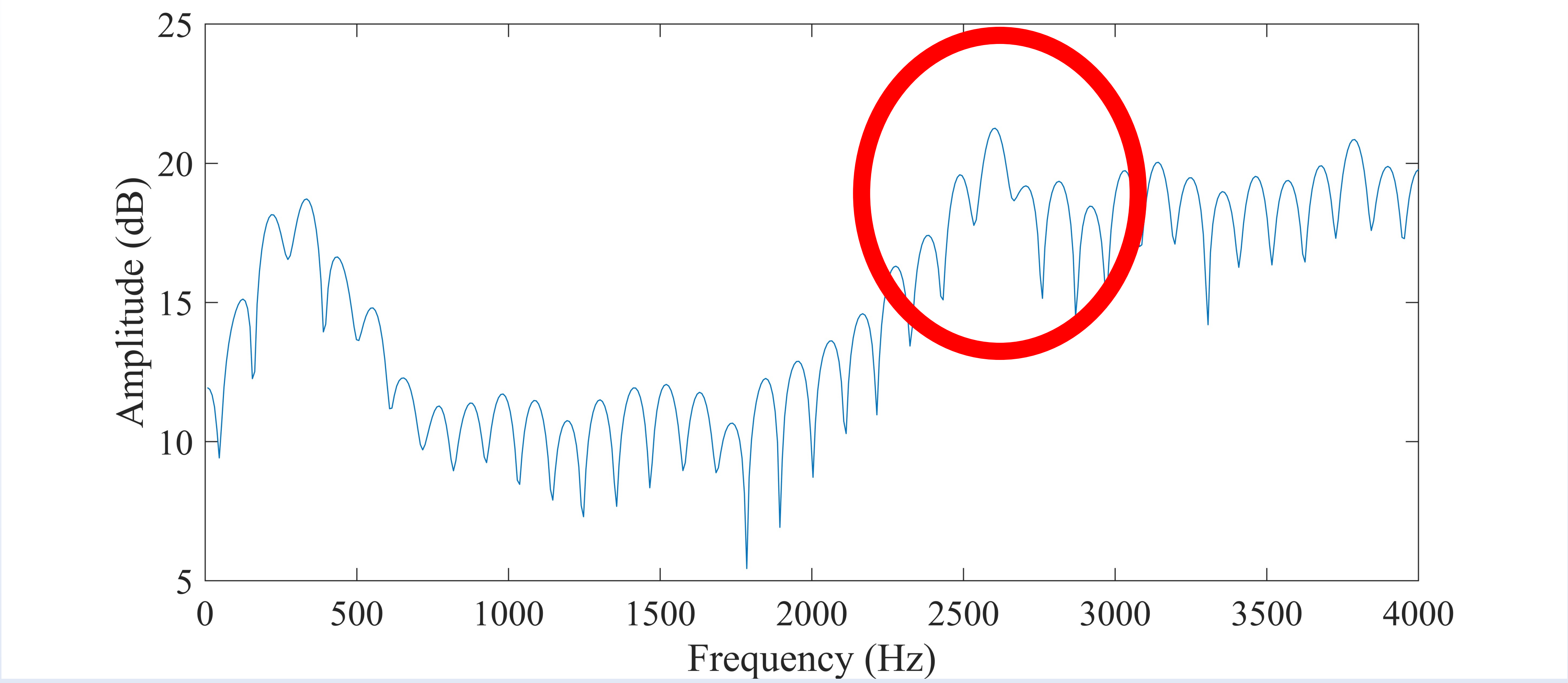
- Collected 55 hours of speech data from 130 K-2nd children
- About a third were bilingual or exposed to a language other than English at home
- Some of the data is longitudinal
- Data included: digits, alphabet sounds, isolated words, GFTA words and sentences, and unscripted narratives

Adult vs. Child Speech Spectra

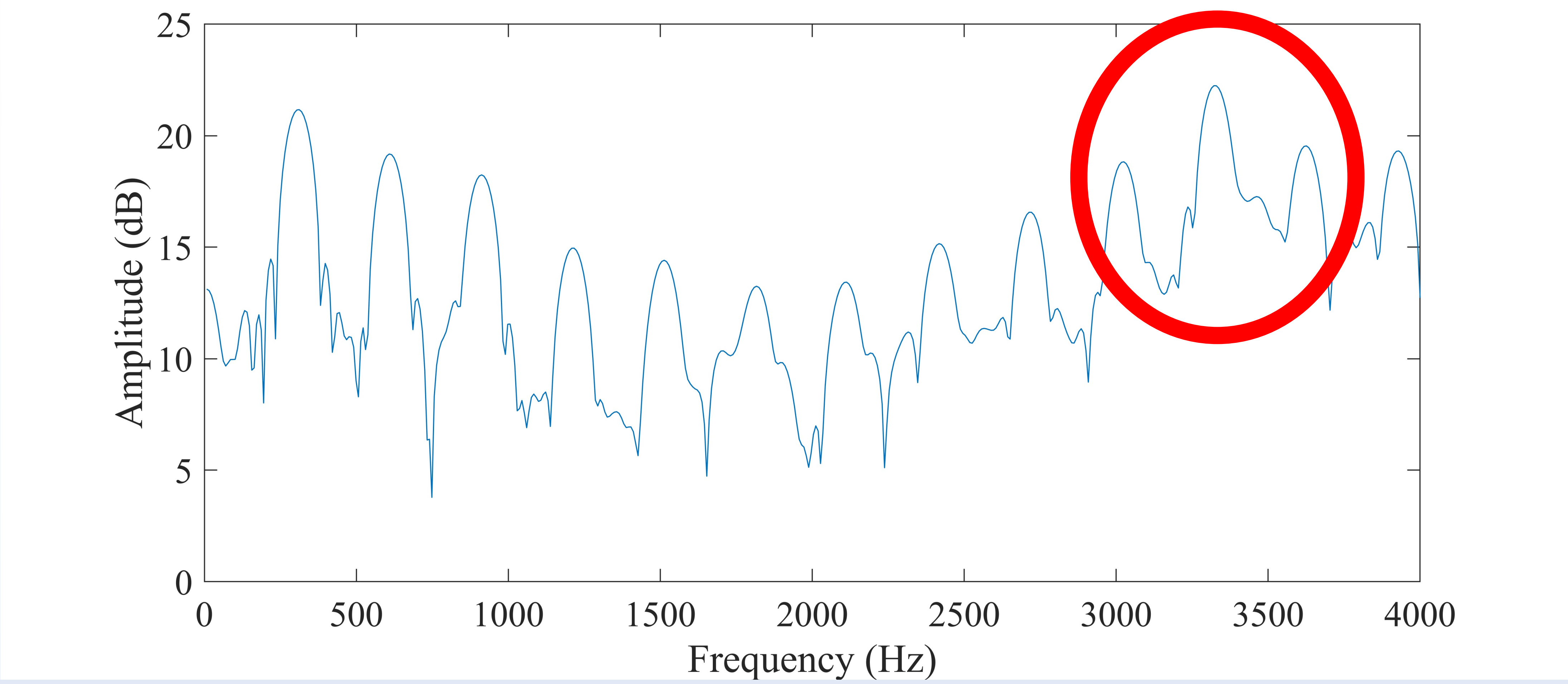




18-year-old male saying /i/

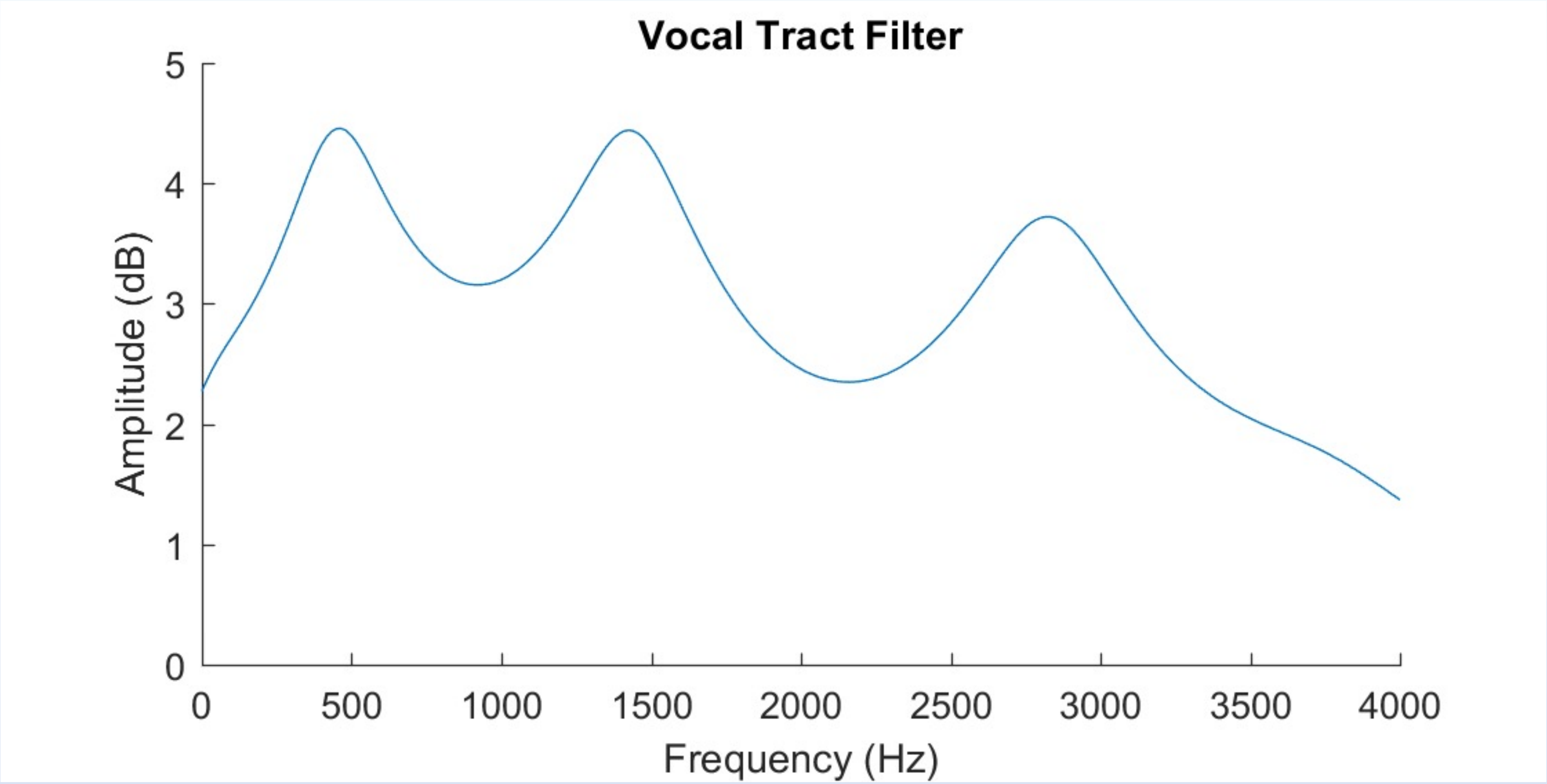


7-year-old male saying /i/



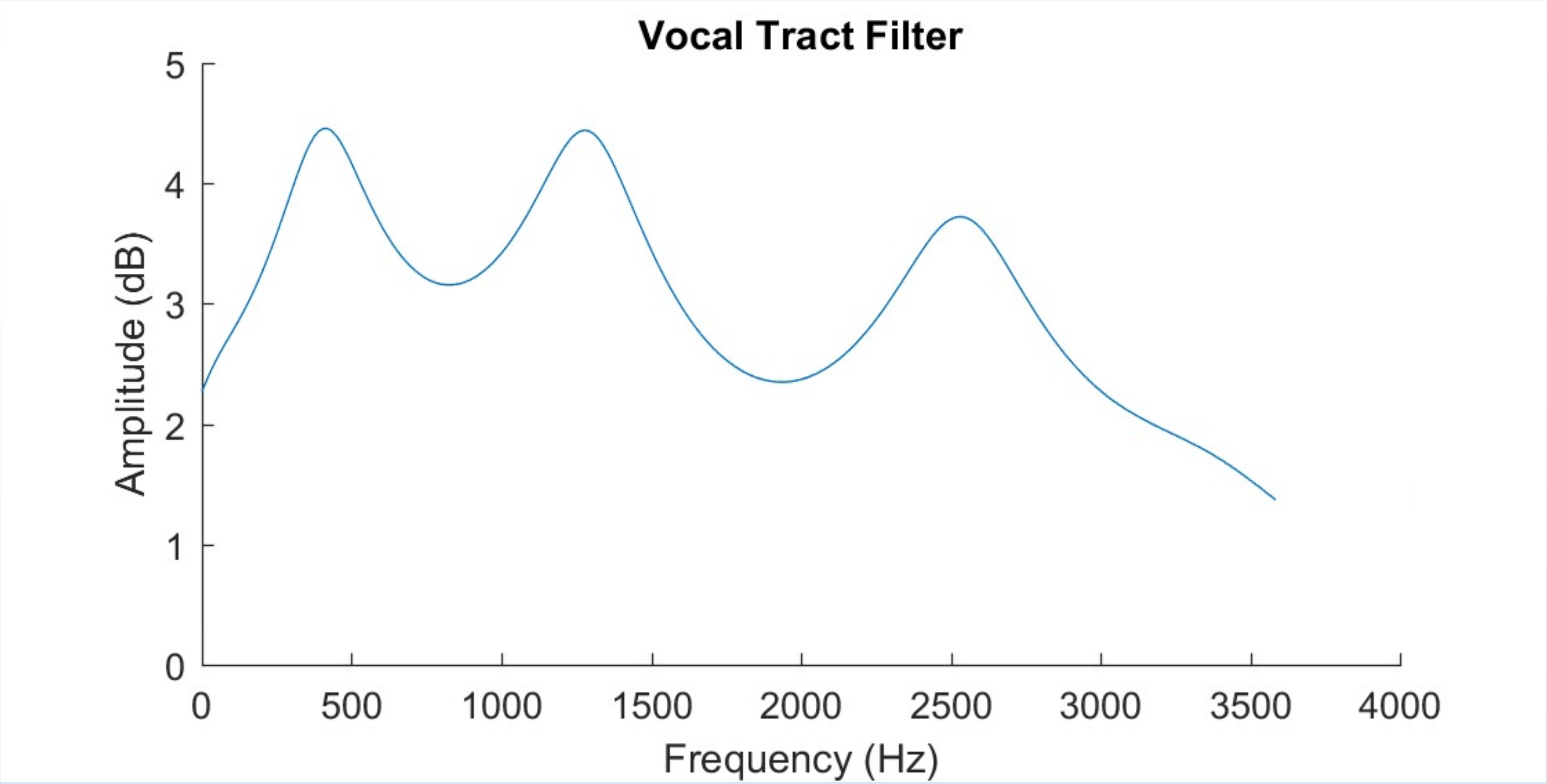
A. Common Technique: Frequency Normalization

- Warp the speech spectra of children to match that of adults or vice versa…



Common Technique: Frequency Normalization

- How to warp?
- Warp the speech spectra of children to match that of adults or vice versa…



Frequency Normalization

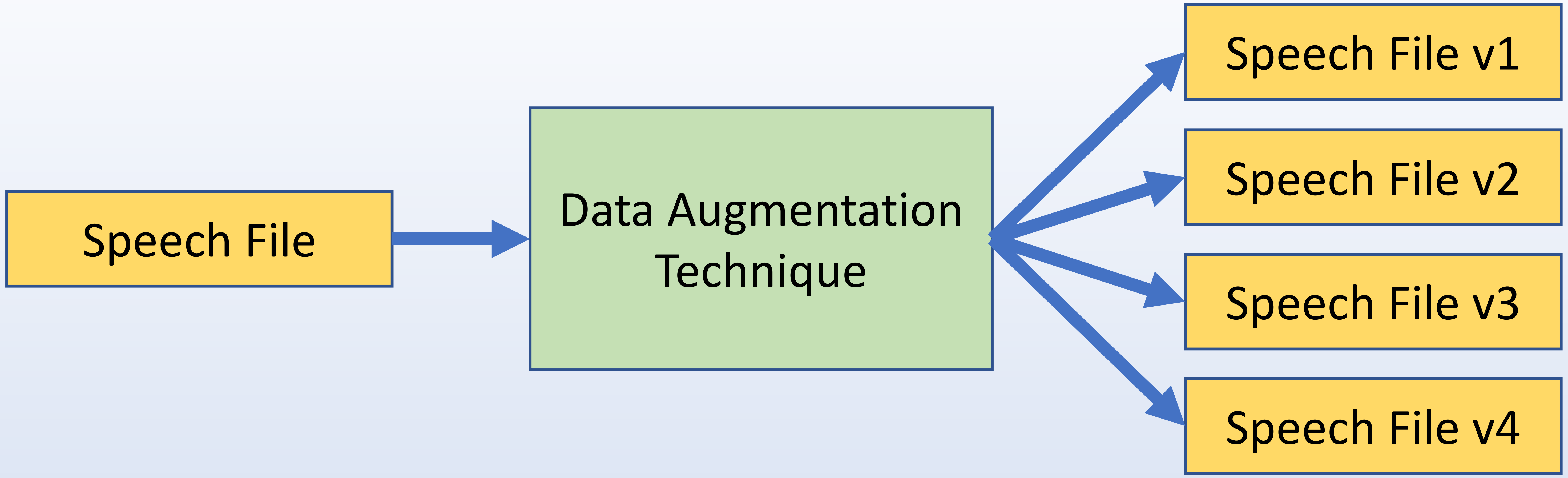
- Using Subglottal resonances as a normalization factor

- Using the fundamental frequency (f_o) as a normalization factor
  - A novel technique for frequency normalization based on models of speech perception for large neural network-based ASR systems and performs better than the widely-used VTLP technique

- Used f_o normalization for data augmentation

B. Data Augmentation

- Neural networks need a lot of data for training and adequate amounts of child speech data are not available
- Extract features in multiple ways from the same data



Continuous Speech ASR

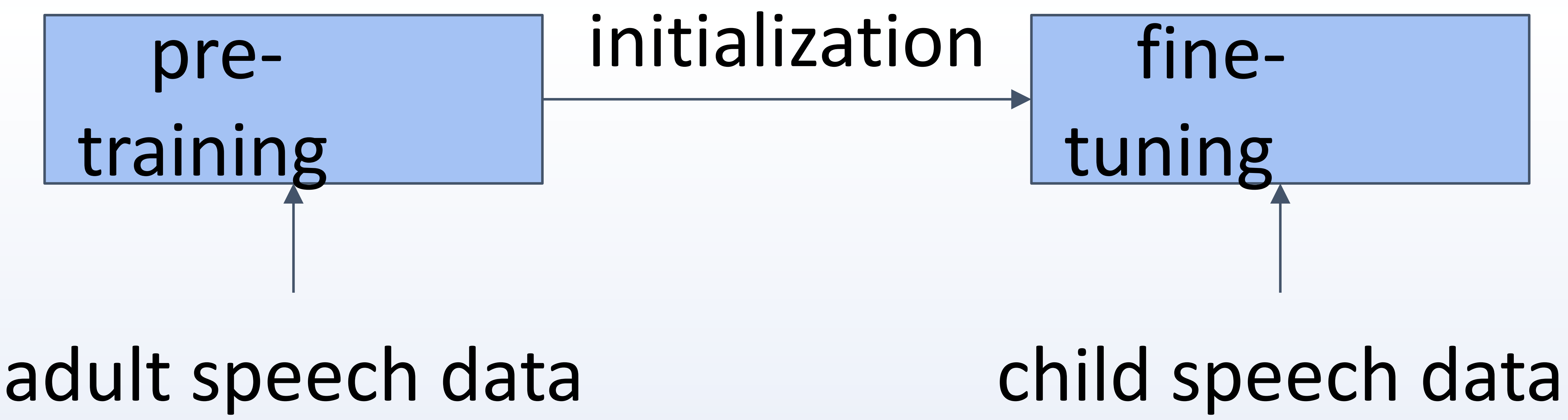
- How do the systems perform per-grade?

WERs (%) of the BLSTM-HMM ASR system for the continuous speech experiment using OGI data. (Yeung et al., 2021)

		Testing Grade					
f_o **Norm?**	f_o **Per?**	K	1	2	3	4	5
No	No	16.97	9.17	6.73	5.71	4.15	4.99
Yes	No	17.44	9.17	6.19	4.80	3.47	4.93
No	Yes	13.97	7.89	5.27	5.11	3.72	4.35
Yes	Yes	**12.87**	**7.38**	**4.88**	**4.78**	**3.35**	**4.24**

C. Our Group is the first to develop Self-Supervised Learning Techniques for Child ASR

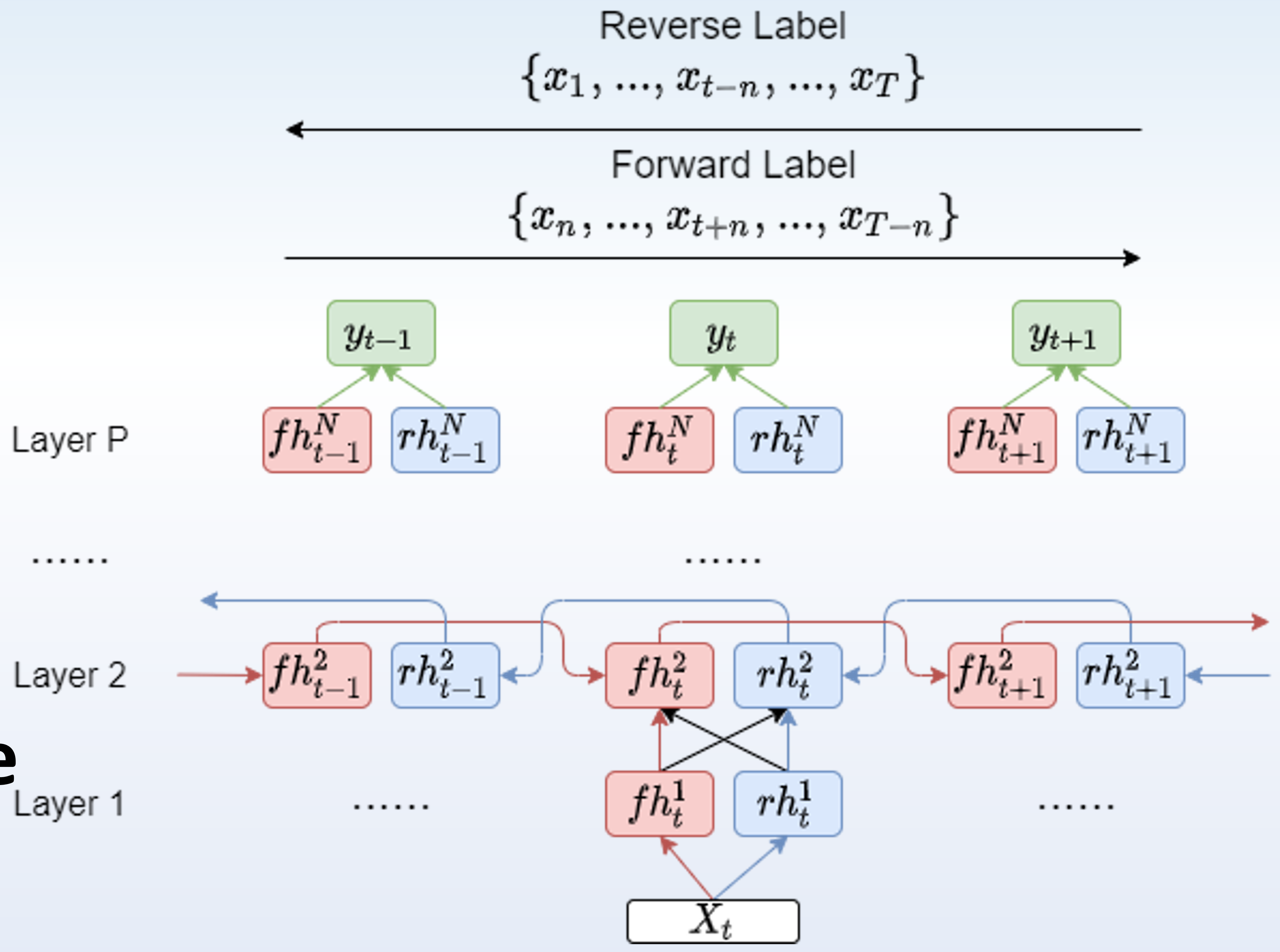
- Two-step process:
    - Pre-training on a data-sufficient task (adult acoustic models)
    - Fine-tuning on the target low-resource task (child acoustic models)



- Pre-training methods depending on whether the pre-training data is labelled:
    - Supervised pre-training (SPT)
    - Unsupervised pre-training (UPT)

Proposed Bidirectional APC (Bi-APC)

- **Proposed Bi-APC:** Decompose forward computation of BLSTM into
  - **Forward path:** predict a frame n steps **after** the current frame given all the **past frames.**
  - **Reversed path:** predict a frame n steps **before** the current frame given all the **future frames.**



Results - Performance on different age groups

BLSTM-based child system performance breakdown based on age groups

WERs(%)	K0-G2	G3-G6	G7-G10
Baseline	18.87	7.24	5.51
+SPT	17.43	6.66	5.11
+APC	18.07	7.03	5.40
+Bi-APC	17.23	6.91	5.26

- Bi-APC provides slightly better results than SPT for younger children
- The larger variability in younger children’s speech causes a large mismatch between pre-training and fine-tuning when using SPT, while Bi-APC can learn more general initial parameters (prior knowledge) for fine-tuning.
- **Current work:** Use Bi-APC for other bidirectional models such as transformers.

Conclusions

Applying linguistic and acoustic knowledge helps to bridge the performance gap in data-driven models for low-resource ASR.

We continue developing techniques for improved Child ASR and currently are studying other dialects such as the African American English dialect.

Publications

- A. Johnson, R. Fan, R. Morris, and A. Alwan, "LPC AUGMENT: An LPC-Based ASR Data Augmentation Algorithm for Low and Zero-Resource Children's Dialects," in ICASSP 2022.
- A. Johnson, A. Martin, M. Quintero, A. Bailey, and A. Alwan, "Can Social Robots Effectively Elicit Curiosity in STEM Topics from K-1 Students During Oral Assessments?" in IEEE Global Engineering Education Conference (Educon) 2022.
- Spaulding, S., Shen, J., Park, H., and Breazeal, C., "Towards transferrable personalized student models in educational games," AAMAS 2021.
- Spaulding, S., Shen, J., Park, H. W., & Breazeal, C. "Lifelong Personalization via Gaussian Process Modeling for Long-Term HRI," *Frontiers in Robotics and AI*, *8*, 152. 2021.
- Gary Yeung, Ruchao Fan, and Abeer Alwan, "Fundamental frequency feature warpingfor frequency normalization and data augmentation in child automatic speech recognition," Speech Communication (2021)
- Jinhan Wang, Yunzheng Zhu, Ruchao Fan, Wei Chu, and Abeer Alwan, "Low Resource German ASR with Untranscribed Data Spoken by Non-native Children – INTERSPEECH 2021 Shared Task SPAPL System," Proc. of Interspeech 2021, pp. 1279-1283
- Ruchao Fan, Wei Chu, Peng Chang, Jing Xiao, and Abeer Alwan, "An Improved Single Step Non-autoregressive Transformer for Automatic Speech Recognition," Proc. Interspeech 2021, pp. 3715-3719
- Ruchao Fan, Amber Afshan, and Abeer Alwan, "BI-APC: Bidirectional autoregressive predictive coding for unsupervised pre-training and its application to children's ASR," ICASSP, 2021, pp. 7023-7027
- Gary Yeung, Ruchao Fan, and Abeer Alwan, "Fundamental frequency feature normalization and data augmentation for child speech recognition," ICASSP, 2021, pp. 6993-6997
- H. Chen, H. W. Park, and C. Breazeal, "Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement," Computers & Education, 2020.

Publications (Cont'd)

- H. Chen, H. W. Park, X. Zhang, and C. Breazeal, "Impact of interaction context on the student affect learning relationship in child-robot interaction," in HRI 2020.
- Trang Tran, Morgan Tinkler, Gary Yeung, Abeer Alwan, and Mari Ostendorf, "Analysis of Disfluency in Children's Speech", Proc. Interspeech 2020, pp. 4278-4282
- I. Grover, H. W. Park, and C. Breazeal, "A semantics-based model for predicting children's vocabulary," in IJCAI 2019.
- Gary Yeung, Alison L. Bailey, Amber Afshan, Morgan Tinkler, Marlen Q. Pérez, Alejandra Martin, Anahit A. Pogossian, Samuel Spaulding, Hae Won Park, Manushaqe Muco, Abeer Alwan and Cynthia Breazeal, "A robotic interface for the administration of language, literacy, and speech pathology assessments for children", SLATE, 2019, pp. 41-42.
- Gary Yeung, and Abeer Alwan "A Frequency Normalization Technique for Kindergarten Speech Recognition Inspired by the Role of F0 in Vowel Perception", Interspeech, 2019, pp. 6-10
- Gary Yeung, Alison L. Bailey, Amber Afshan, Marlen Q. Pérez, Alejandra Martin, Samuel Spaulding, Hae Won Park, Abeer Alwan and Cynthia Breazeal "Towards the Development of Personalized Learning Companion Robots for Early Speech and Language Assessment", AERA, 2019.
- S. Spaulding, H. Chen, S. Ali, M. Kulinski, and C. Breazeal, "A social robot system for modeling children's word pronunciation: Socially interactive agents track," AAMAS 2018.
- Gary Yeung, Steven M. Lulich, Jinxi Guo, Mitchell S. Sommers, and Abeer Alwan "Subglottal resonances of American English speaking children", The Journal of the Acoustical Society of America 144 (6), 3437-3449, 2018
- Gary Yeung and Abeer Alwan "On the Difficulties of Automatic Speech Recoginition for Kindergarten-Aged Children", in Proc. Interspeech 2018, pp. 1661-1665