Development, Deployment and Evaluation of Personalized Learning Companion Robots for Early Literacy and Language Learning (NSF 1734380)

MIT: Cynthia Breazeal, Hae Won Park
UCLA: Abeer Alwan; Alison Bailey
UW: Mari Ostendorf

https://robot-literacy-tools.media.mit.edu/

This NRI project will develop, deploy and evaluate **personalized companion robots** to assist kindergarten-age children in learning language and vocabulary skills. The aim is to accelerate the impact of social robots for **early childhood education** in schools and at home. The project will generate new insights for how to develop expressive, socially responsive robots that provide more effective, engaging, and empathetic educational experiences for young children.

Long-term Personalized Reading Companion

Key Problem and Significance

Personalizing to Student's Learning
- Personalized intervention is the most effective method for early literacy and language learning but it is hard to achieve in classroom settings

Personalizing Robot Behavior to Maximize Engagement
- Each student learns and is motivated differently. The robot should learn an interaction policy to maximize each student's engagement.

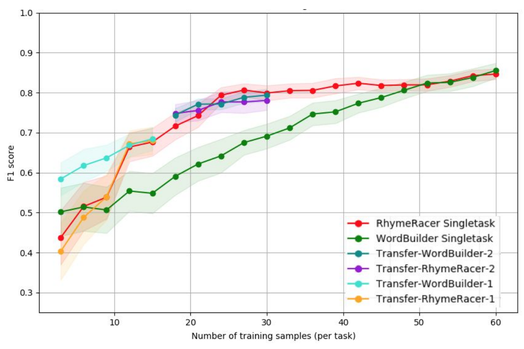
Reliable, Robust, Affordable Long-term Robot Platform
- Demonstrate and evaluate a long-term deployable robot system in-school and at-home.

Scientific Impact

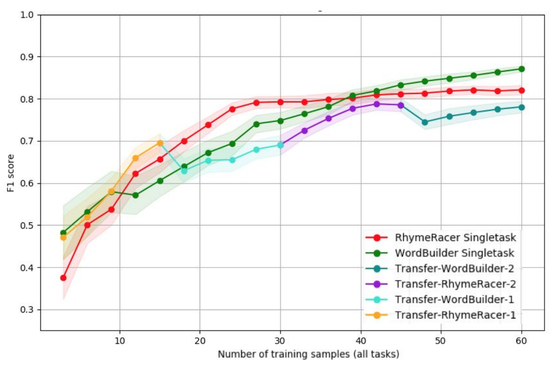
- Limited data and narrowly-focused interactions are two challenges in developing personalized models. **Learning across related tasks** can significantly improve the accuracy and data efficiency of student model learning [1].
- Engagement detection is crucial for personalized learning. Students show contextually variant facial affective cues when learning new information. **Understanding these affective cues based on the context** of the interaction can guide robot's action policy to maximize learning outcomes [2,3].

Solutions (Approaches, Innovations, Contributions)

- **Multitask Personalization** a paradigm for designing student models that are **transferrable** across tasks, improving model data efficiency and domain flexibility [1]
- **Gaussian Process-based model** set in a joint word-space domain gives two game tasks a shared representation [1]
- **Instance Weighting protocol** transfers prior data based on task-similarity w.r.t observed data points [1]



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



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

- Robot approximates **optimal role-switching policy** while **incorporating facial affective engagement cues** to maximize learning gains.
- Rewards from affective features are **conditioned on the context** in which behaviors are exhibited [3].
- Student's knowledge is estimated using Bayesian inference and further incorporated into the learning algorithm.

Broader Impact

- **Engaging Teachers and Parents: Outreach, Guidelines, and Best Practices** - 1) Focus group sessions with all stakeholders to discuss experiences with implementation of the social robots in school/home settings. 2) Review of third-party video of assessment sessions during which student level of engagement and language/literacy performance in different settings and with different partners can be evaluated. From these efforts we anticipate isolating best practices & establishing guidelines that we'll freely distribute.
- **Datasets, Tools, Education Apps**: a set of databases, long-term robot platform tools, and educational activities developed for use with the social robot platform will be disseminated to the research and education community.
- **Cross-Disciplinary Training for Students**: cross-disciplinary training through Annual Workshops including topics: reliability and validity of technology-based systems, child development and early education, performance assessments, advanced user interfaces and technologies, experimental design and analysis, spoken language technologies, etc.

[1] Spaulding, S., Shen, J., Park, H., and Breazeal, C. "Towards transferrable personalized student models in educational games," in AAMAS 2021.

[2] 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.

[3] 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.

[4] G. Yeung, R. Fan, and A. Alwan, "Fundamental frequency feature normalization and data augmentation for child speech recognition", to appear in ICASSP, 2021.

[5] R. Fan, A. Afshan, and A. Alwan, "BI-APC: Bidirectional autoregressive predictive coding for unsupervised pre-training and its application to children's ASR", to appear in ICASSP, 2021.

[6] T. Tran, M. Tinkler, G. Yeung, A. Alwan, and M. Ostendorf, "Analysis of Disfluency in Children's Speech", Interspeech 2020.

Dialogic QnA

Key Problem and Significance

- In dialogic reading, the companion asks questions to confirm the child's understanding and maintain engagement
- Questions should be timed based on child engagement and turntaking cues
- Flexible timing requires contextually grounded question generation

Scientific Impact

- Attending to prosodic cues improves decisions on **question timing**
- Contextually grounded **question generation** enables child-sensitive timing of questions
- **Novel corpora** of disfluency annotated child speech and questions support work on child-oriented conversational AI

Solutions (Approaches, Innovations, Contributions)

- **Question timing**
  - *Challenge:* lack of research on child conversational speech
  - Analyze adult-child interactions to identify prosodic cues in child speech that trigger adult turntaking vs. backchannels.
  - Explore child disfluencies as an indicator of uncertainty.
  - Learn prosodic cues in neural net for deciding the next dialog act.

- Contextually grounded automatic **question generation**
  - *Challenge:* lack of data for learning text generation
  - Leverage question-answering corpora (inverting the task: given answer, predict question) and a pre-trained model (BART)
  - Mix with small set of QA pairs for children's stories
  - Multi-dimensional automatic & crowdsourced human evaluations
- **Experiment Results:**
  - Manual ratings show near-human performance for child-appropriateness & context relevance, but a few grammatical errors

Automatic Speech Recognition (ASR) for Children

Key Problem and Significance

- Current state-of-the-art ASR systems perform poorly for children's speech and thus do not properly assess children's oral language abilities due to, in part, the lack of child speech databases to train and develop effective child ASR systems.
- Effective ASR systems would lead to personalized educational tools.

Scientific Impact

- Creation of a novel longitudinal children's speech database.
- Development of new supervised and unsupervised training techniques for robust ASR systems.
- Development of novel transfer learning and data augmentation techniques for low resource ASR tasks.

Solutions (Approaches, Innovations, Contributions)

- **Database Creation**
  Creation of longitudinal children's speech database that documents how each child's voice develops annually
- **Novel techniques for low-resource ASR tasks**
  Development of methods in transfer learning, domain adaptation, and data augmentation that will be used to improve children's ASR systems and can be applied to other low resource domains
  - Learning prior knowledge from untranscribed adult speech as initialization to aid in child ASR training.
  - Augmenting child speech data based on fundamental frequency normalization without changing the transcription.

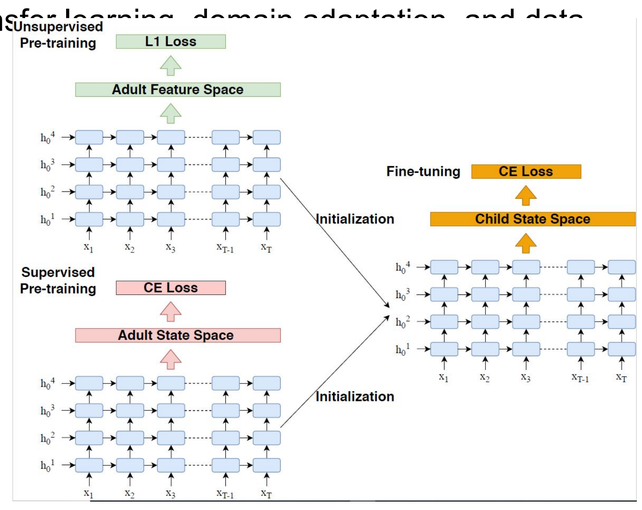


Illustration of supervised and unsupervised pre-training (transfer learning). The blue parts are the LSTM layers to be transferred.

- **Children's Speaker Identification (SID) Systems**
  Development of children's SID systems in order to automatically identify child speakers and personalize the robot's interactions with each child leading to personalized learning, diagnostics, and assessment.