Context Adaptations for Improving Child Automatic Speech Recognition

Manoj Kumar[1], Daniel Bone[1], Kelly McWilliams[2], Shanna Williams[2], Thomas D. Lyon[2] and Shrikanth Narayanan[1]

[1]Signal Analysis and Interpretation Lab, USC
[2]USC Child Interviewing Lab

Motivation

- ASR forms a crucial component in **automatic understanding of spoken language**
- Child ASR is a **harder problem** than Adult ASR (Lee et al '99)
- **Behavioral cues** from adult speech are **indicative of child mental state**
  - Paralinguistic behavior & Natural language use provide cues of ASD severity (Bone '14, Kumar '16)
  - Prosodic features vary significantly with child's engagement levels (Gupta '16)

*We hypothesize that **context helps in speech recognition**, during **both automatic** speech recognition and **human** speech recognition*

Objective

Understanding child spoken behavior during adult-child interactions

- Improving child automatic speech recognition (ASR)
- Borrow information from the **context of interaction**
- **Understand type & effect of context** on ASR

Forensic Interviews

- **Child maltreatment:** One of the most serious threats to children's well-being (Norman '12, Fang '12)
- Need for **systematic interview format** with both recall and recognition questions.
- **Objective methods** to train attorneys w.r.t linguistic and paralinguistic behavior

Datasets

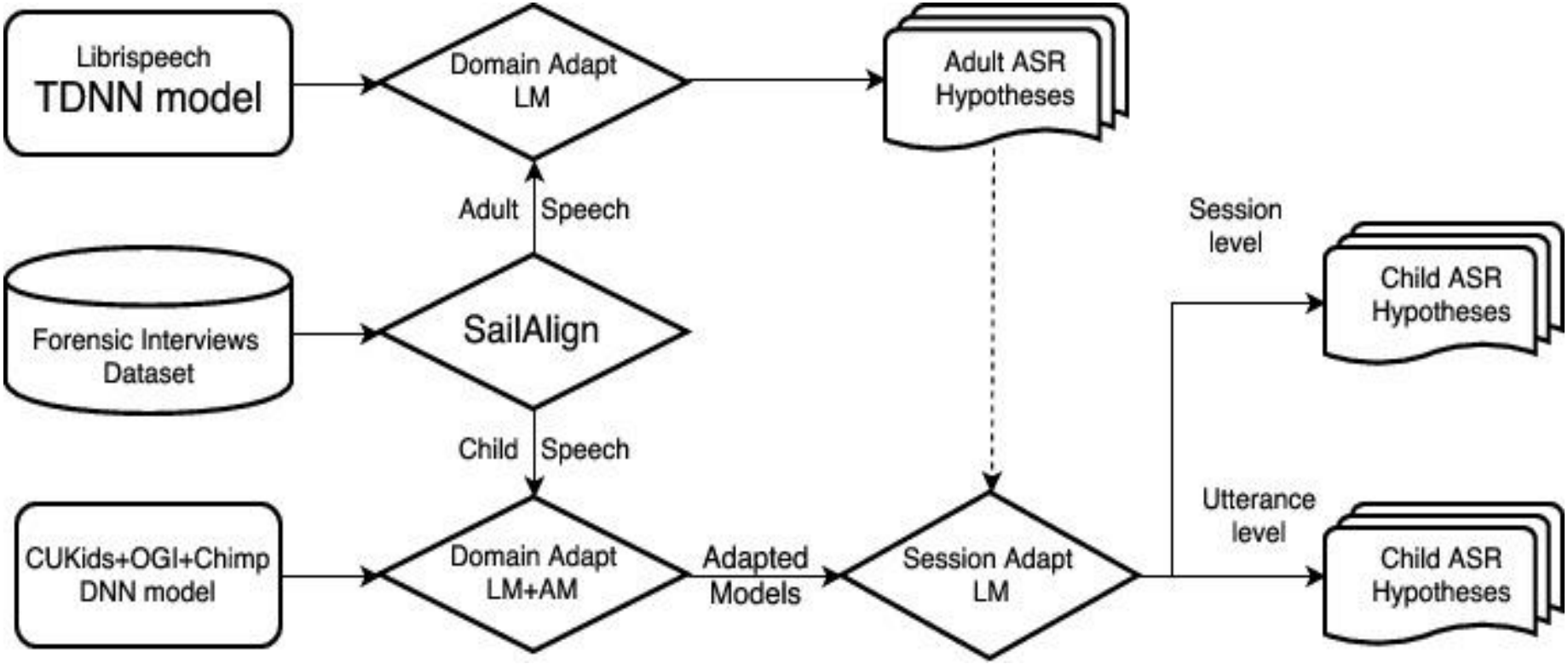
Training Corpora

- **CUKids:** Read speech including isolated words, sentence & stories
- **CHIMP:** Spontaneous speech data collected while playing video game
- **OGI:** Prompted & spontaneous speech including alphabets and sentences
- **Librispeech:** Read speech corpora of audio books

Evaluation Corpus

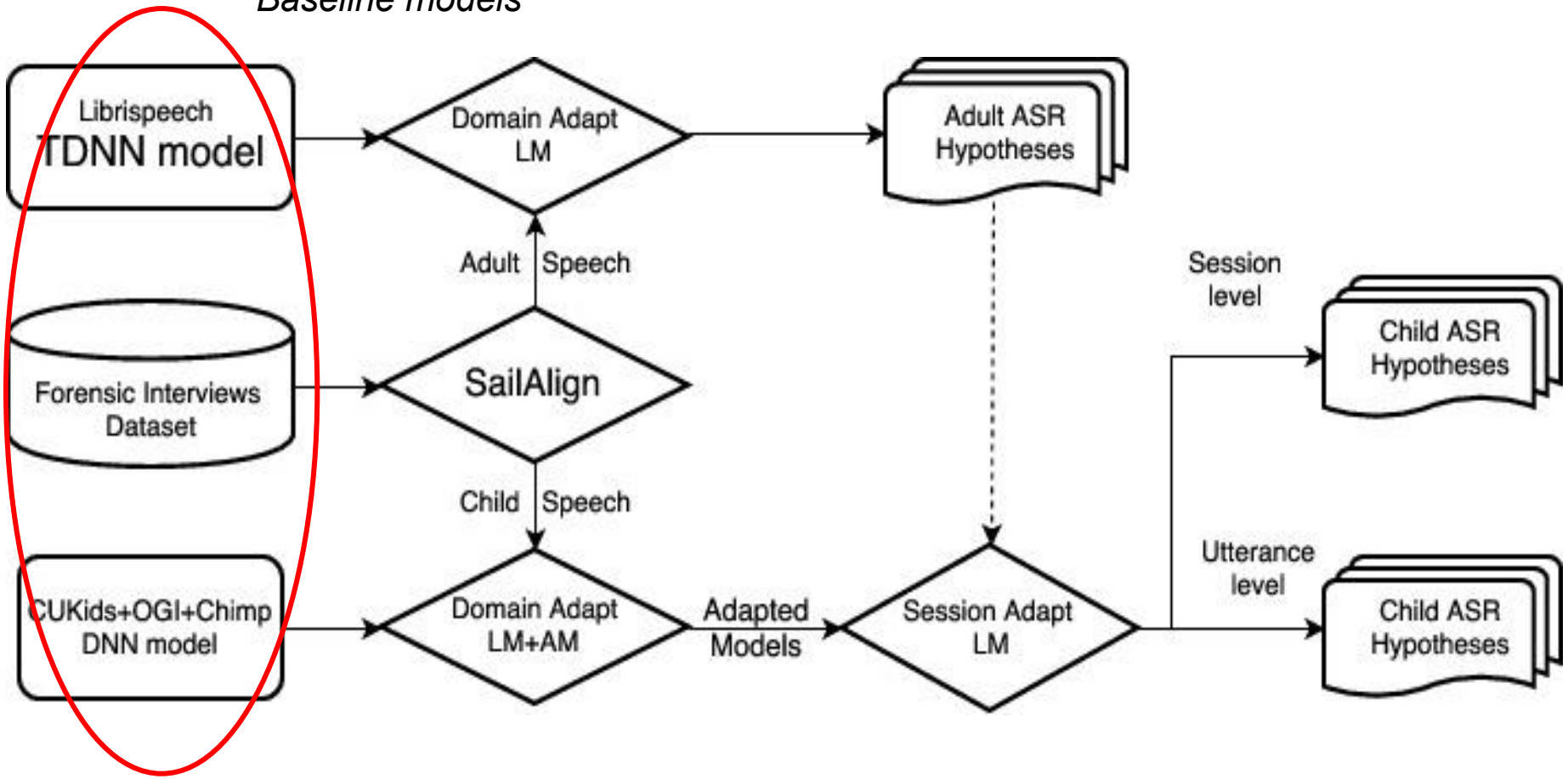
- **Forensic Interviews:** Spontaneous conversational speech
  - 30 children from age group 4 -12 yrs

Methodology



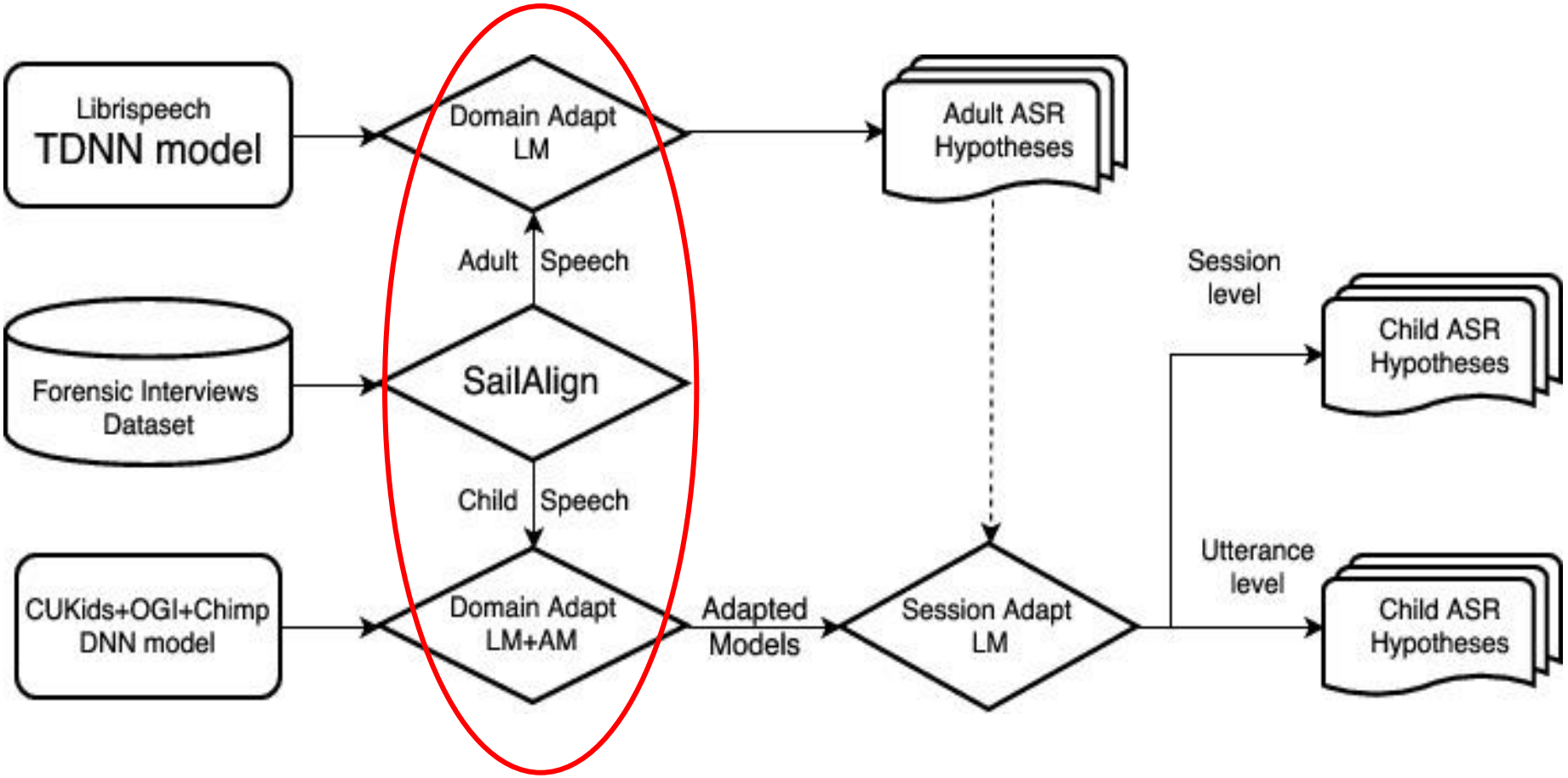
Methodology

Data preparation &
Baseline models



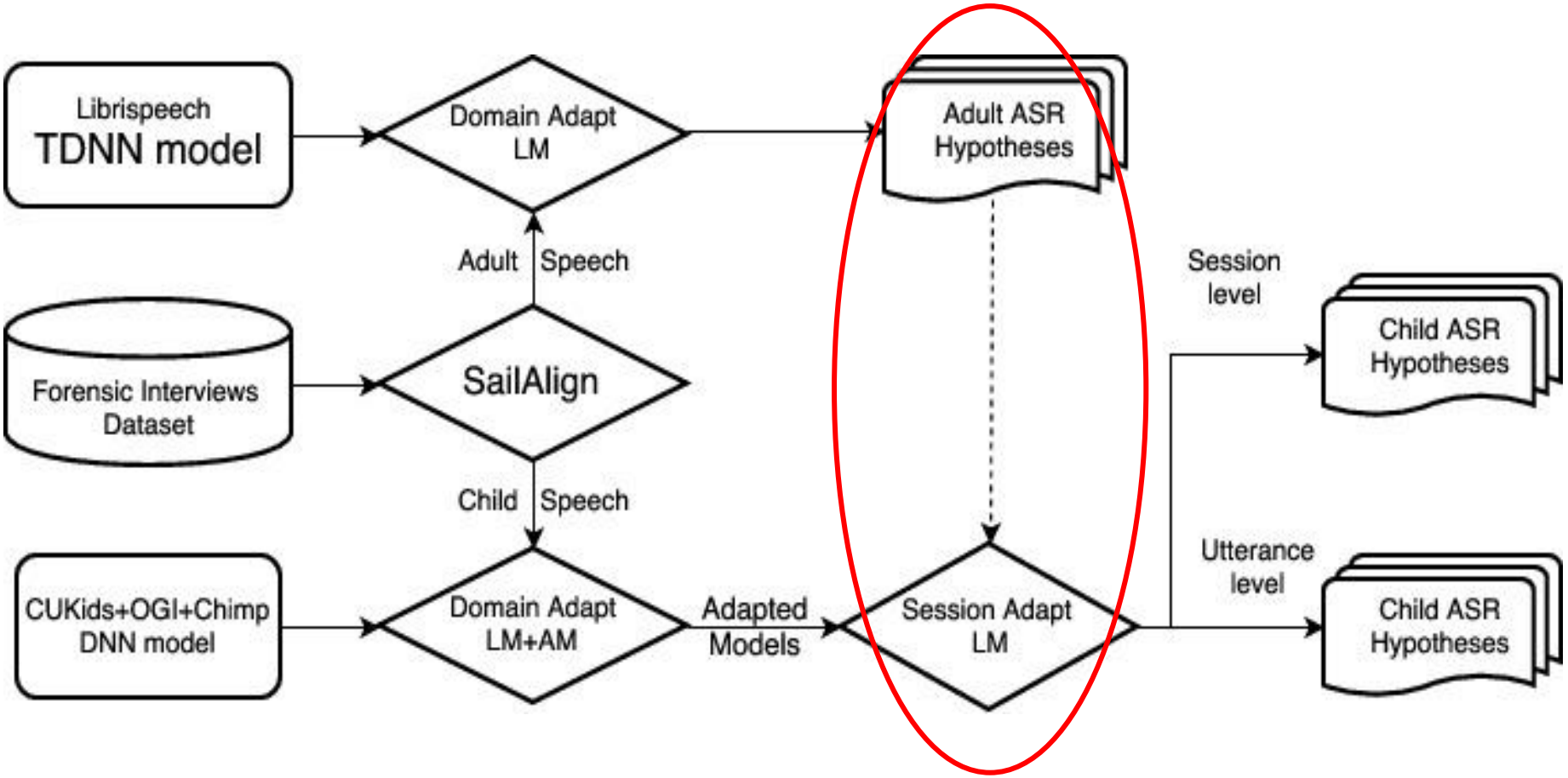
Methodology

Domain adaptations



Methodology

Session adaptation



Adaptation Results I

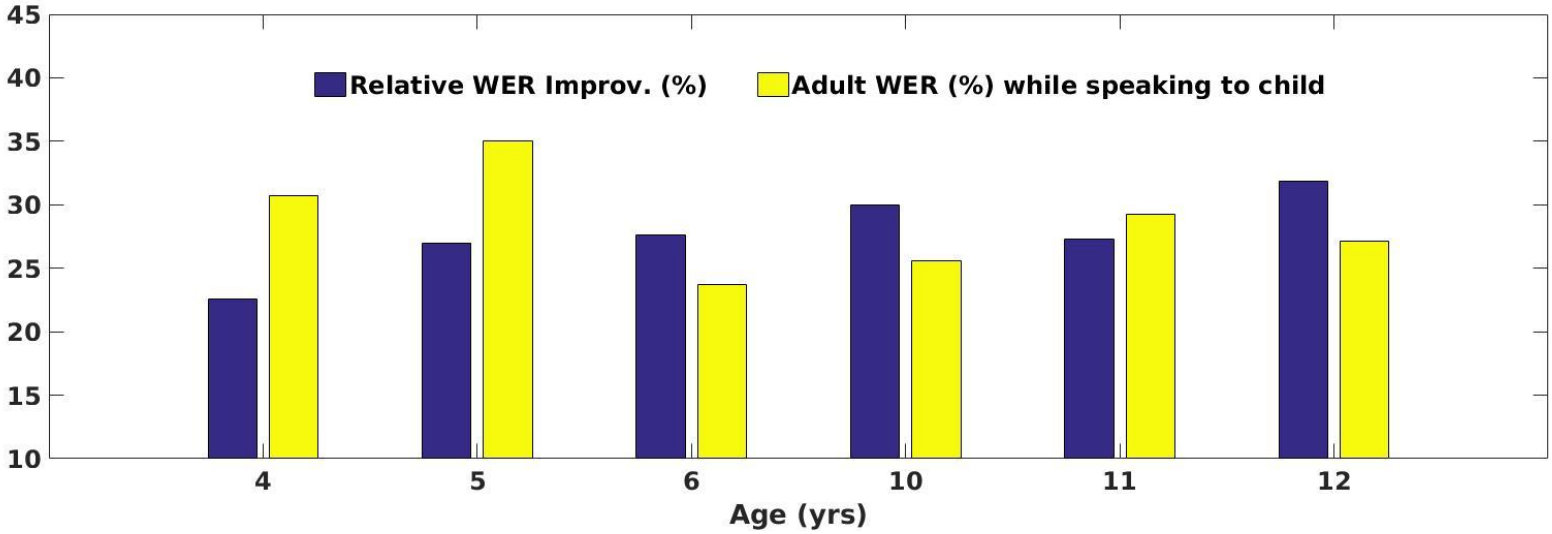
	Adult Speech	Child Speech	WER(%)	Mean Perplexity
Baseline	Data: Librispeech ***(100hrs)*** AM: 7-layer TDNN LM: Tri-gram	Data: CUKids+CHIMP+OGI ***(80hrs)*** AM: 7-layer DNN LM: Tri-gram	73.39	431
Domain adapt. ***(3.7 hrs)***	LM: Linear interpolation	LM: Linear Interpolation AM: Re-train baseline with adaptation data	62.47	247
Session adapt. ***(~8min)***	-	LM: Linear interpolation	61.04 *(Global)* **52.69** *(Local)*	207 **193**

Significant improvements in WER and perplexity over baseline

Adpatation Results II

Do context-based benefits vary by age?

We analyze overall improvement in session adaptation over baseline



Larger improvements for older children possibly due to ***more accurate adult speech recognition***

Adaptation Results III

Does direction of context affect child ASR?

We repeat session-adaptation by conditioning on the number of context utterances

# Utts	1	2	3	4
Forward	53.98	54.30	53.58	54.40
Backward	50.78	49.41	49.15	***47.47***
Combined	52.69	50.80	49.66	49.55

- Forward direction independent of context size - ***It does not matter how far we listen ahead***
- Backward direction improves with context size - ***Interviewer borrows word-counts from child speech***

HSR (Listening) Experiment

Does context aid in human speech recognition?

We asked 3 native English speakers to transcribe the test data under two conditions:

- ***Without context*** By listening to only current utterance ***(WER:27.08)***
- ***With context*** Listening to previous & following utterance; followed by current utterance ***(WER: 22.49)***

Adult: OK AND WHO DOES <*name*> SHARE A ROOM WITH?
Child: NOBODY JUST COOKIE MONSTER
Adult: JUST COOKIE MONSTER OK AND WHO DOES <*name*> SHARE A ROOM WITH?

20.4% relative WER improvement *with inclusion of context*

Conclusions

Conclusion

- Conditioning on the interlocutor's speech improves child ASR
- Session-level adaptation beneficial **only when localized**

Future Work

- **Incorporate semantic information** during LM adaptation - topic models, dialogue completion systems (Serban '15)
- Automatically **select context** for adaptation

Acknowledgement

This work was supported by funding from National Institutes of Health (NIH) and National Science Foundation (NSF)

Thank You!