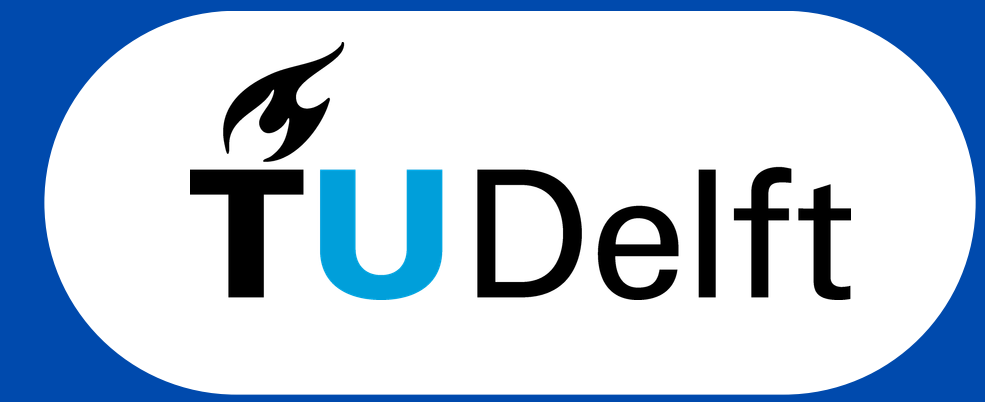


Utterance Length and ASR Performance Disparities Between Native and Non-Native Dutch-Speaking Children

Student: Zev Comvalius | Responsible Professor: Odette Scharenborg | Supervisor: Yuanyuan Zhang
EEMCS, Delft University of Technology



1. Introduction

- **Automatic Speech Recognition (ASR)** systems are increasingly used in speech-based services, but they perform worse for **children and non-native** speakers.
- This project studies whether **utterance length** partly explains the ASR performance gap between native and non-native Dutch-speaking children.
- Short utterances give ASR systems less context and make Word Error Rate (WER) more sensitive to individual errors.

2. Research Questions

Main:
To what extent is utterance length associated with ASR performance disparities between native and non-native Dutch-speaking children, as measured by word error rate, and how does this differ between read speech and HMI speech?

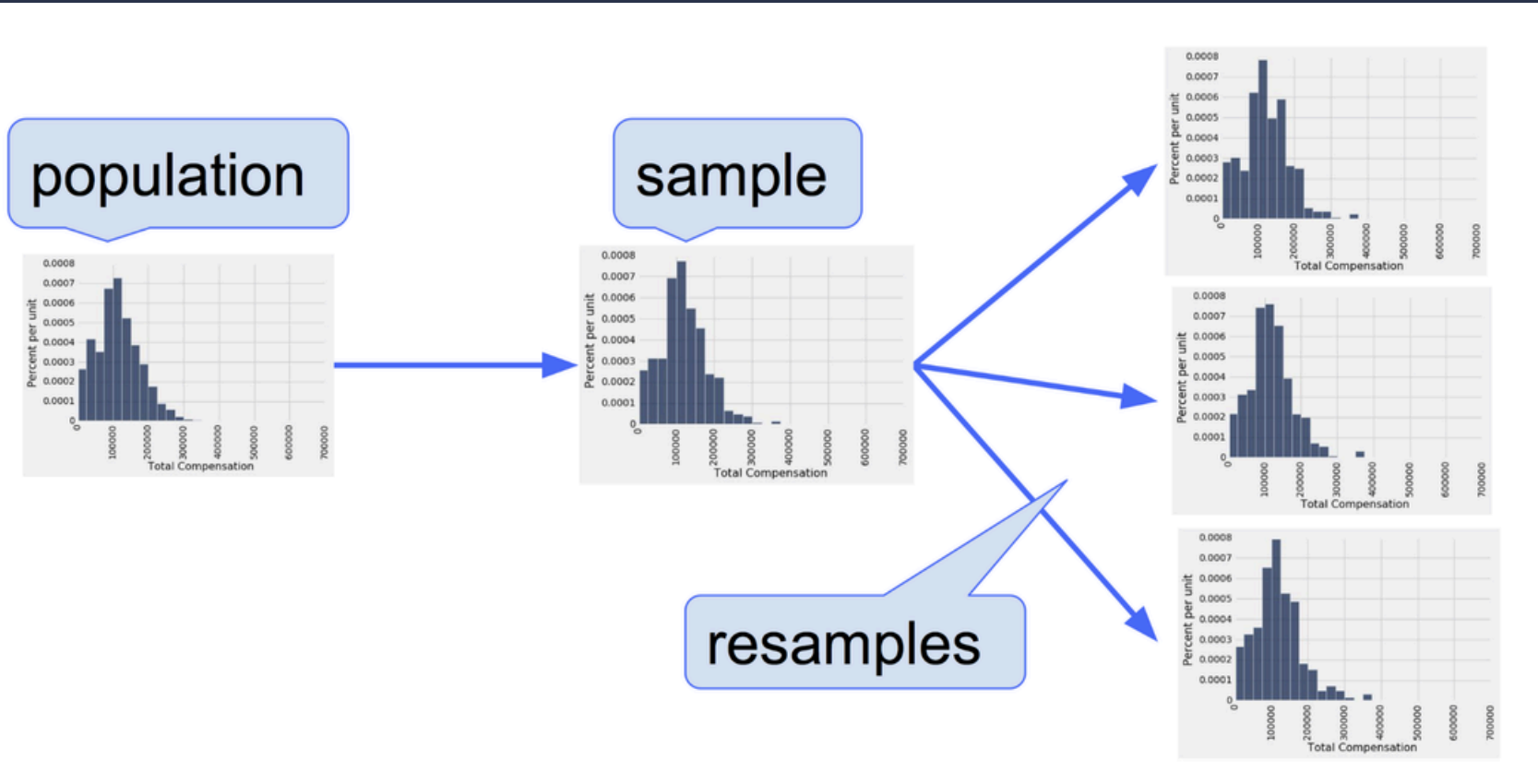
- Sub:**
- Native vs. non-native distribution of utterance lengths
 - Google Chirp 2 vs. Whisper-large-v3
 - Word Error Rate and native/non-native WER gap

3. Data & ASR Systems

- **Corpus:** Dutch child speech from the Jasmin corpus
- **ASR Systems:** Google chirp and Whisper-large-v3
- **ASR outputs:** Outputs of ASR systems with corpus utterances as inputs

4. Methodology

- **Utterance length** = words in reference text.
- **WER** = errors / reference words
- **Performance disparity** = WER non-native - WER native
- The analysis compared:
 1. Overall native/non-native WER gap
 2. WER across utterance lengths
 3. Group differences in length distributions
 4. Gaps after length standardization
 5. Patterns across both ASR systems
- Uncertainty was estimated using **speaker-level bootstrap** resampling.



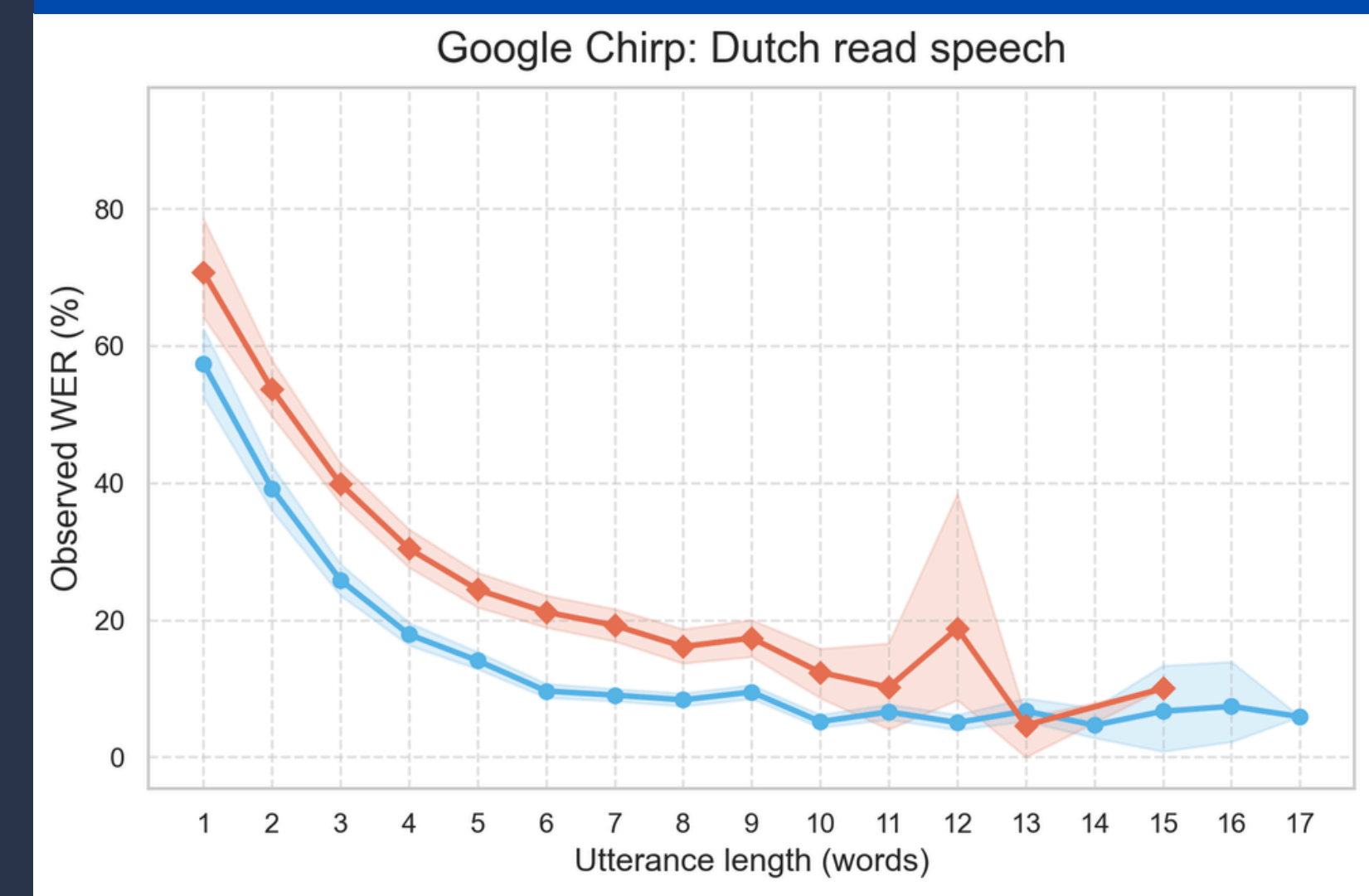
5. Results

- **Baseline gap:** Non-native WER is higher in every condition, and the gap ranges from 12.9 to 16.5 pp
- Read speech **gap slope:** -1.9 pp/word (Chirp), -2.1 pp/word (Whisper)
- Standardization reduced read gaps by about 3 pp

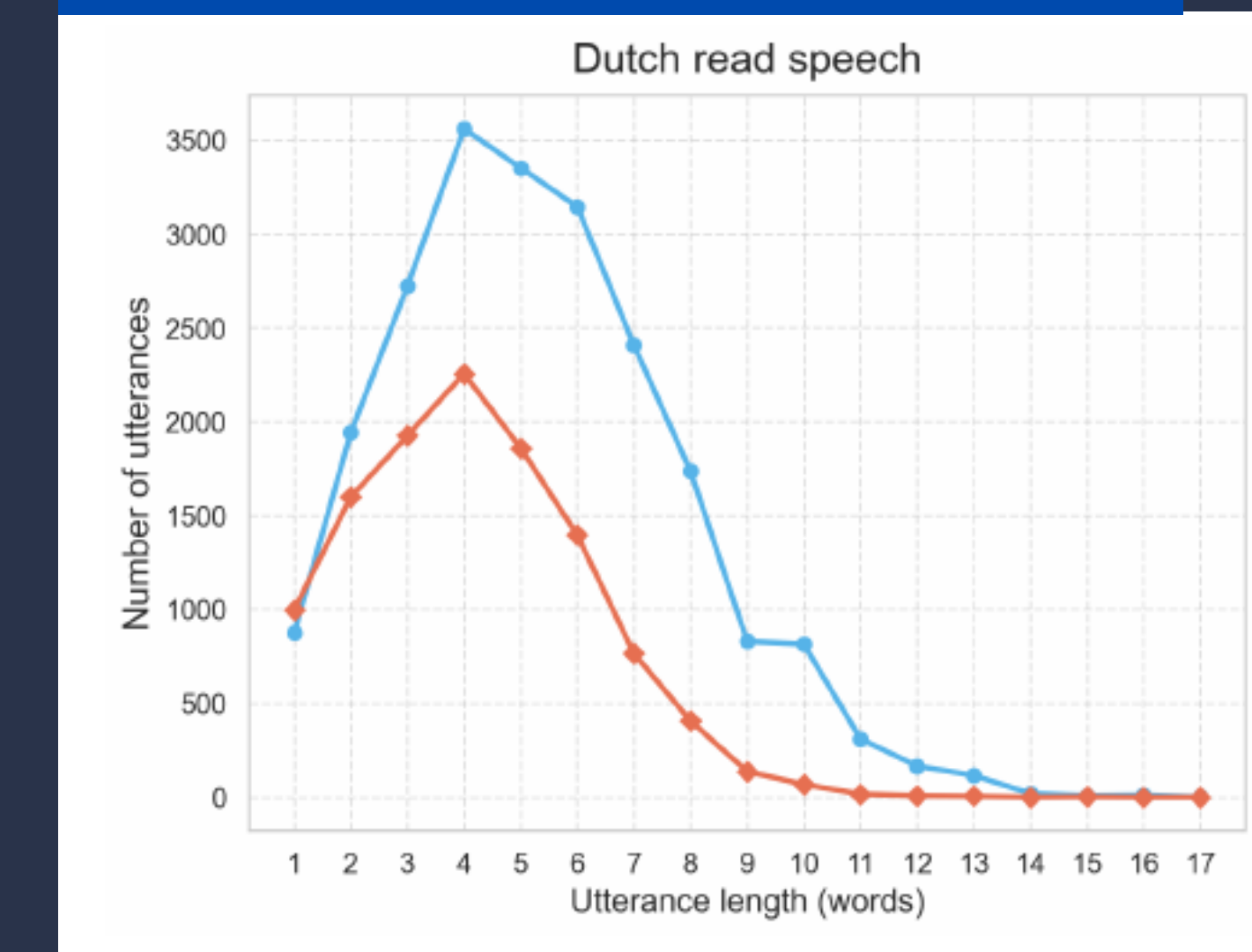
Result: Baseline gap

ASR	Speech	Native	Non-native	Gap
Chirp	Read	13.0%	28.2%	+15.2 pp
Chirp	Hmi	21.0%	33.9%	+12.9 pp
Whisper	Read	14.5%	31.1%	+16.5 pp
Whisper	Hmi	25.3%	40.1%	+14.8 pp

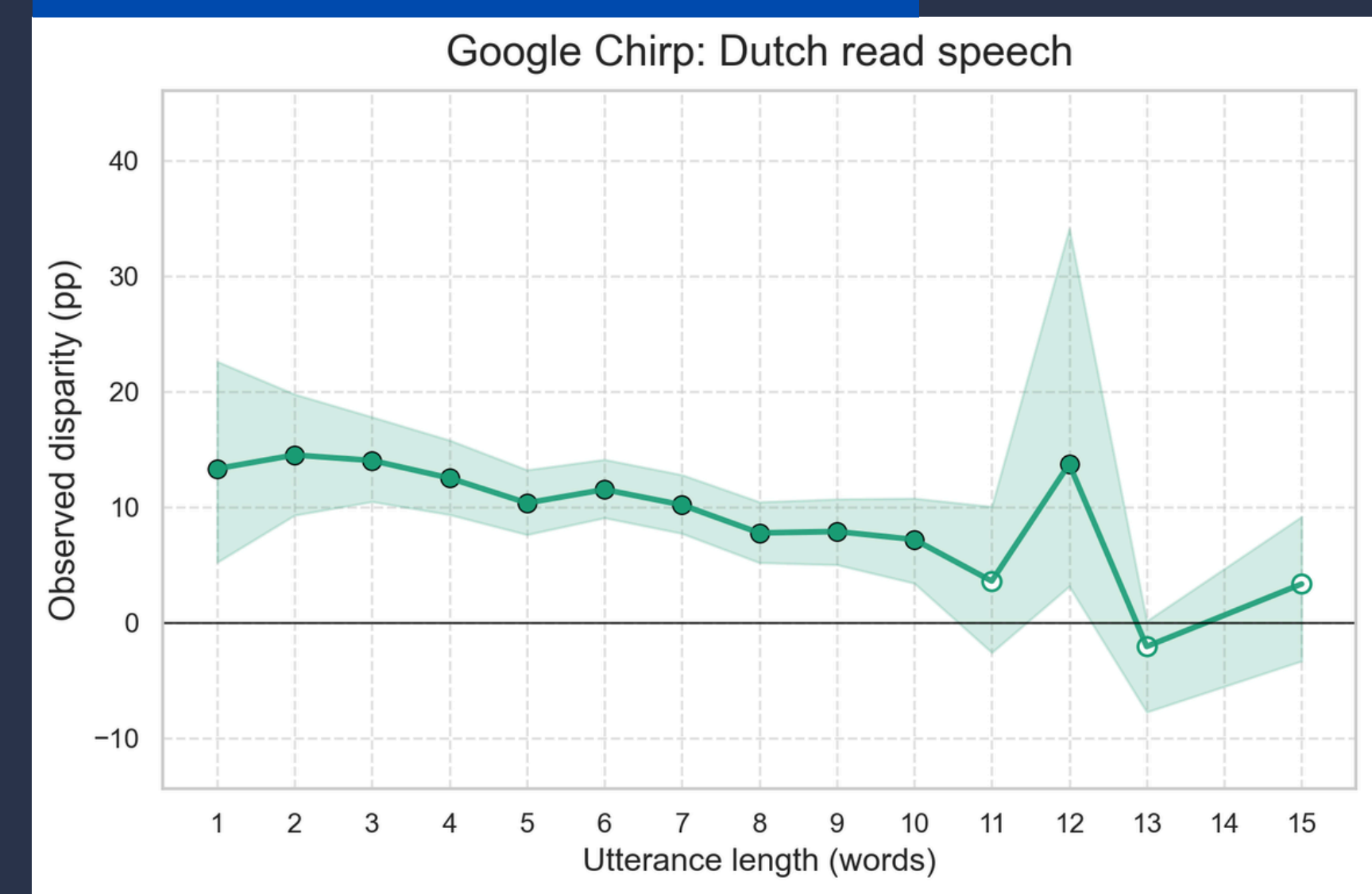
Result: WER decreases with length



Result: Length distributions



Result: Disparity by length



6. Conclusion

- Utterance length is associated with ASR performance disparities!
- **Main finding:** Shorter utterances have higher WER, and the native/non-native gap is larger for short utterances.
- Length explains part of the read-speech disparity, but it does not fully explain the disadvantage for non-native child speech.