

UC Davis

UC Davis Previously Published Works

Title

Comparing intelligibility and recognition memory of human and text-to-speech voices

Permalink

<https://escholarship.org/uc/item/98w818qv>

Journal

The Journal of the Acoustical Society of America, 153(3_supplement)

ISSN

0001-4966

Authors

Aoki, Nicholas
Zellou, Georgia

Publication Date

2023-03-01

DOI

10.1121/10.0018532

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Comparing intelligibility and recognition memory of human and text-to-speech voices

Nicholas B. Aoki and Georgia Zellou
University of California, Davis

INTRODUCTION

Clear speech improves both intelligibility and recognition memory (Van Engen et al., 2012)

Effortfulness Hypothesis: Reduced perceptual effort (e.g., clear speech) leaves more resources available in working memory (McCoy et al., 2005)

Advances in TTS synthesis => TTS speaking styles varying in intelligibility (Aoki et al., 2022).

Little work has examined effects of TTS styles on memory for listeners from diverse backgrounds

Current Study and Predictions

We test effects of voice type (human, TTS), speaking style (clear, casual), and listener age (younger, older) on intelligibility and memory.

Intelligibility: Human > TTS (Aoki et al., 2022); Clear > Casual (Cohn et al., 2021); Younger > Older (Kim et al., 2006)

Memory:

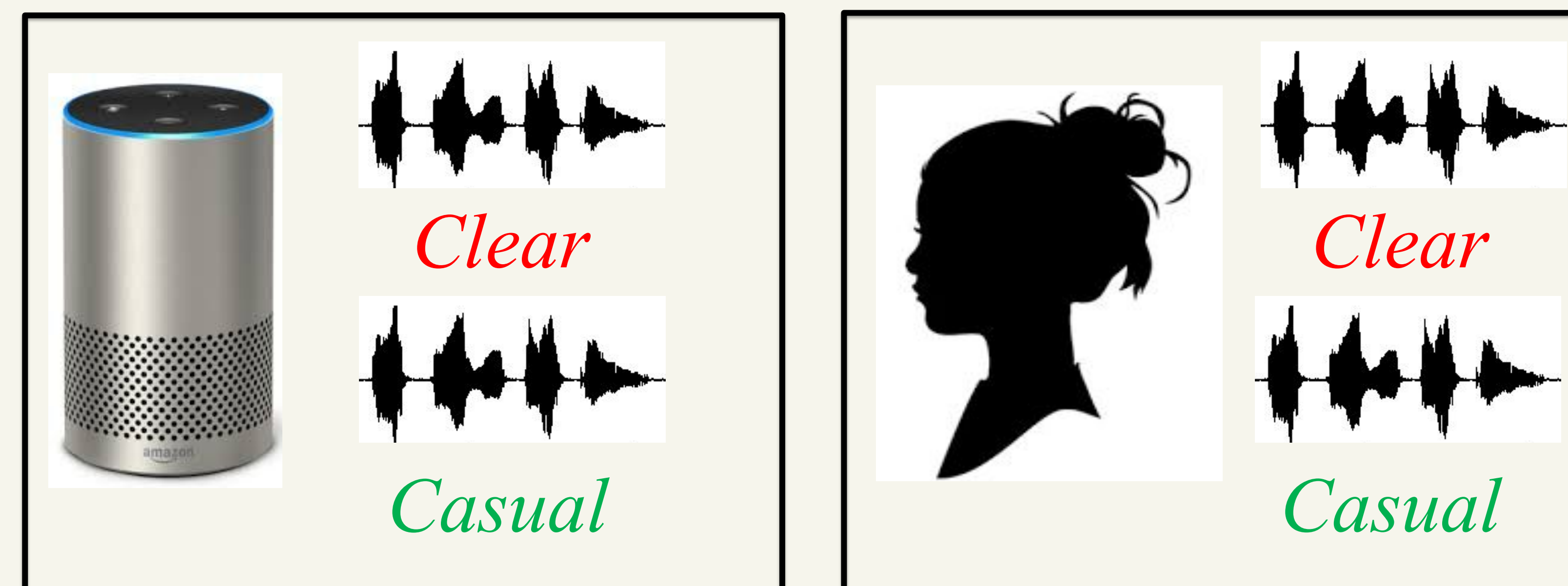
- Effortfulness Hypothesis: Effects on memory should match intelligibility
- TTS “clear” and “casual” speech differ acoustically from naturally produced clear speech => potentially different effects on memory
- Older listeners may be more adversely impacted by casual speech and TTS voices

METHODS

Native English speakers recruited on Prolific

Experiment 1: Intelligibility

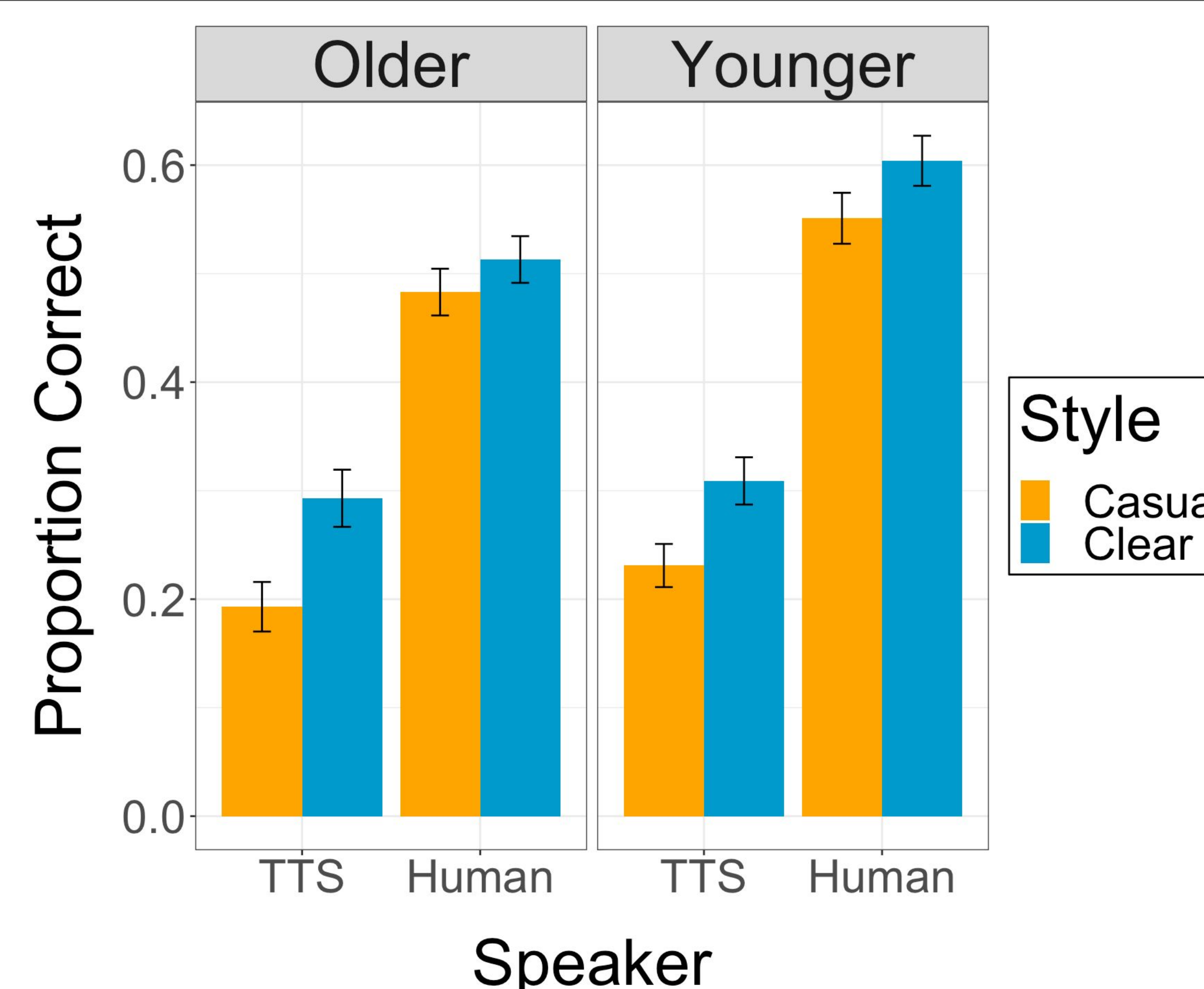
- Younger (n = 30; 18-30) and older (n = 30; 50+)
- Speech-perception-in-noise task
- Voice = between subjects; Style = within subjects



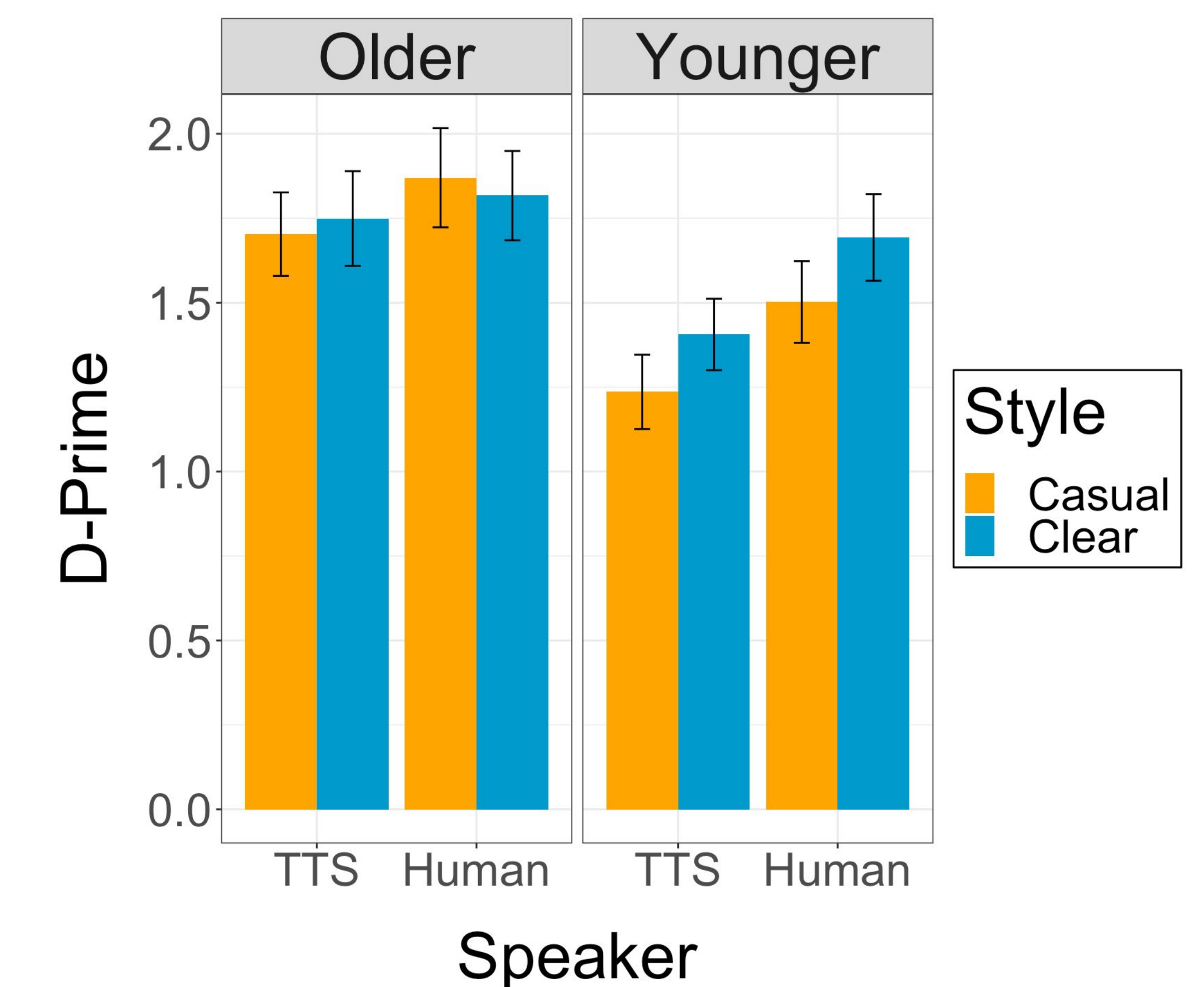
Experiment 2: Recognition Memory

- Younger (n = 60; 18-30) and older (n = 66; 50+)
- Part I: Passively listened to 30 sentences
- Part II: Heard 60 sentences, indicated whether the sentence was old (i.e., heard before) or new
- Voice = between subjects; Style = within subjects

RESULTS: INTELLIGIBILITY



RESULTS: MEMORY



CONCLUSION

Discussion

Intelligibility and memory results matched on Voice (TTS voices = lower intelligibility and memory), but not Style or Age

- Effect size of intelligibility may be crucial (Van Engen et al. speaker = +15% accuracy in clear compared to casual; current human speaker = ~4% difference; TTS speaker = ~7% difference)
- Older listeners = greater effort?

Future Directions

- Examine how individual differences in clear/casual speech affect memory
- Effects of noise on recognition memory
- Understanding how synthetic speech can be better remembered has important implications for TTS voice development