Effects of voice continuity and stimulus rate on auditory working memory

Sung-Joo Lim[1,2,3], Barbara G. Shinn-Cunningham[2,3] & Tyler K. Perrachione[1,3]

[1] Department of Speech, Language & Hearing Sciences [2] Biomedical Engineering
[3] Center for Research in Sensory Communication & Emerging Neural Technology, Boston University

Introduction

- Speech processing can be cognitively demanding, especially when the acoustics of incoming speech signals have high variability [1].
- The human auditory system perceptually adapts to speech characteristics of a talker (i.e., talker adaptation), which facilitates rapid and accurate speech recognition [2, 3].
- However, it is unknown (i) whether the facilitatory effects of talker adaptation persist beyond immediate speech recognition, and (ii) how the effects change when listeners have more time to process individual speech tokens.
- Research Questions:
  - Does talker adaptation lead to perceptual benefits in speech encoding or recall of speech information maintained in working memory?
  - Do the facilitatory effects from talker adaptation change with the amount of time to process incoming speech?

Methods

Participants

- 27 young adults with normal hearing (age: 18–30 years)

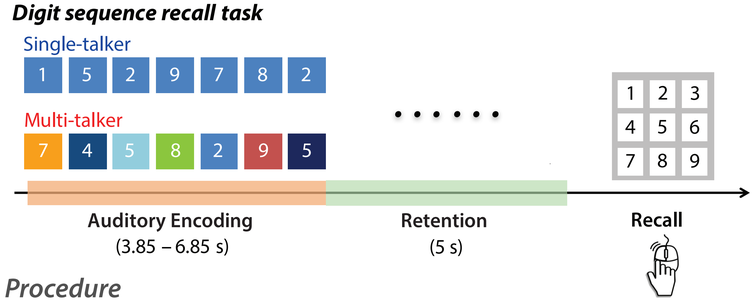
Stimuli

- Naturally-spoken digits (1–9), recorded by eight native American-English talkers (4 female; 4 male)
- Durations of all recordings were normalized to 550 ms (Praat ver. 5.3).

Task design

- Encode a sequence of seven, randomly selected digits, and recall the sequence in the order of presentation after a 5-s delay.
- Design (2 *talker* × 3 *stimulus rate)*:
  - *Talker* conditions: digits spoken by one single talker vs. multiple talkers
  - *Stimulus rate* conditions: 0-, 200-, and 500-ms inter-stimulus intervals (ISIs) between the digits in the sequence during encoding

Digit sequence recall task



Procedure

- 24 trials per block (× 6 blocks); blocked stimulus rate conditions; semi-randomized talker conditions (3 consecutive trials of a single talker or multiple talkers)

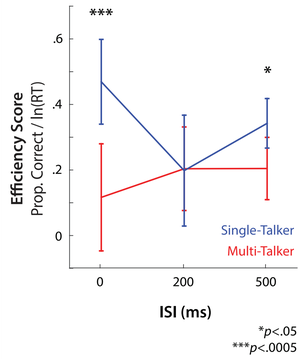
Behavioral data

- Proportion of correct recall; log-transformed onset response time
- Efficiency score = proportion correct / ln (onset RT)

Preliminary EEG experiment:

- EEG recording & analyses: a 64-channel active EEG system (BioSemi) sampling at 2000 Hz; average mastoids reference; 1–100 Hz filtered

Voice continuity leads to efficient recall



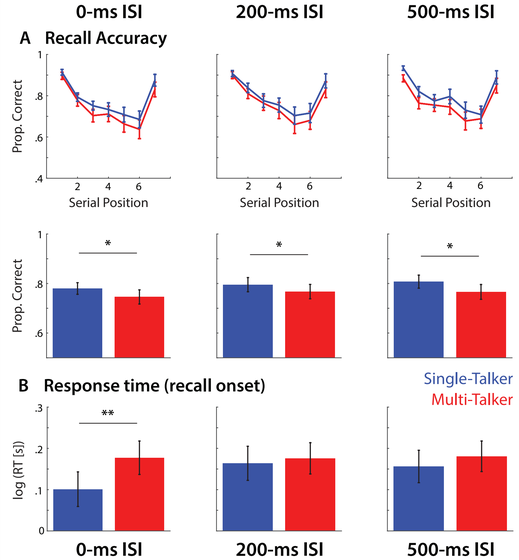
*p<.05
***p<.0005

Figure 1. Efficiency of digit sequence recall. Recall efficiency was computed as the average accuracy of sequence recall weighted by onset response speed (log-transformed inverse response time). Higher recall efficiency indicates more accurate and faster recall of digit sequences. For the faster presentation rate (0-ms ISI), listeners were significantly more efficient (i.e., faster and more accurate) in recalling sequences spoken by a single talker than multiple talkers.

Talker: $\chi^2(1) = 12.80$, $p = .00035$***
ISI: $\chi^2(2) = 0.58$, $p = .75$ (n.s.)
Talker × ISI: $\chi^2(2) = 9.90$, $p = .007$ **

ISI effect in the Single-talker: $\chi^2(2) = 9.80$, $p = .007$ **
ISI effect in the Multi-talker: $\chi^2(2) = 1.51$, $p = .47$ (n.s.)

Performance benefits from voice continuity



Talker: $F(1,26) = 13.65$, $p = .001$**
ISI: $F(1.9,50.3) = 1.56$, $p = .22$
Position: $F(1.7,44.7) = 27.5$, $p < .0005$***
Talker × ISI: $F(1.9,49.5) = .28$, $p = .75$
Talker × Position: $F(4.6,119) = .52$, $p = .75$
ISI × Position: $F(7.9,204.5) = 1.94$, $p = .06$
Talker × ISI × Position: $F(7.5,194) = .74$, $p = .65$

Talker: $F(1,26) = 13.65$, $p = .001$**
ISI: $F(1.9,50.3) = 1.56$, $p = .22$
Talker × ISI: $F(1.9,49.5) = .28$, $p = .75$

Talker: $F(1,26) = 7.47$, $p = .011$*
ISI: $F(1.6,41) = .99$, $p = .36$
Talker × ISI: $F(1.9,49.1) = 2.0$, $p = .15$

*p<.05
**p<.005
***p<.0005

Figure 2. Digit sequence recall performance in 2 (*talkers*) × 3 (*stimulus rate; ISI*) conditions. (A) Response recall accuracy. (Top) average recall performance across participants (N=27) as a function of digit position in the sequence. (Bottom) average recall performance across digit positions. Across all ISIs, participants exhibited consistently higher accuracy in recalling digit sequences spoken by a single talker than multiple talkers. **(B)** Response onset time. At the fastest presentation rate (0-ms ISI), listeners responded faster for sequences spoken by a single talker vs. multiple talkers. Error bars: ± 1 SEM.

References

[1] Nusbaum, H. C. & Magnuson, J. S. (1997). *Talker variability in speech processing*, pp. 109–132.
[2] Mullennix, J. W. & Pisoni, D. B. (1990). *Perception & Psychophysics, 47*(4): 379–390.
[3] Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2017). *Attention, Perception, & Psychophysics.*
[4] Jensen, O. & Mazaheri, A. (2011). *Frontiers in Human Neuroscience, 4*: 186.
[5] Shinn-Cunningham, B. G. (2008). *Trends in Cognitive Sciences, 12*(5): 182–186.

Neurophysiological signatures

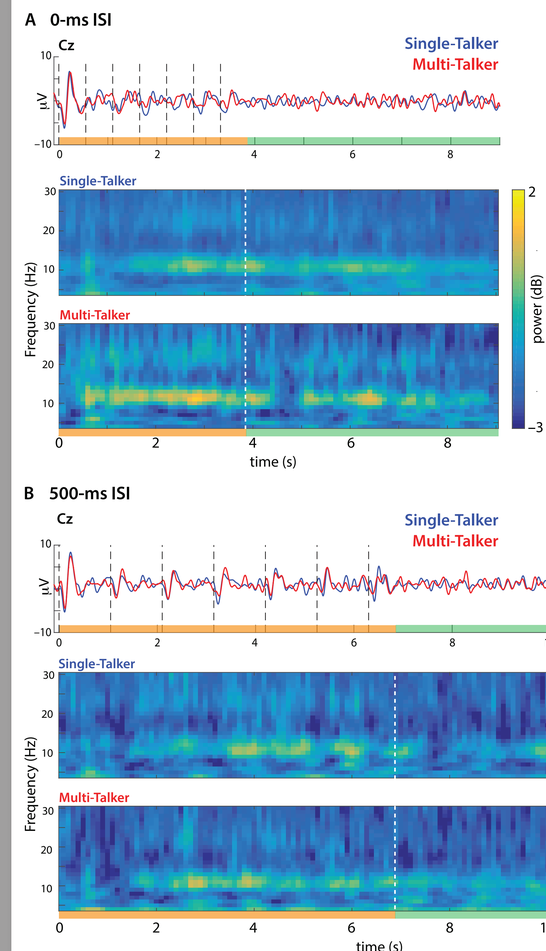


Figure 3. Preliminary EEG responses throughout the trial. The EEG experiment tested two stimulus presentation rates (0- and 500-ms ISIs). Grand average neural responses for the faster **(A)** and slower **(B)** presentation rates. (Top panels) evoked response potentials relative to the trial onset in the single vs. multiple talker conditions. (Bottom panels) time-frequency representations during task performance in the single vs. mulitple talker condition. Modulations of neural oscillatory power in the alpha range (~10 Hz) [4] seem to be greater at the faster presentation rate (A) during both stimulus encoding and memory retention phases.

Discussion

- Talker adaptation enhances accuracy and leads to faster recall during working memory for speech.
- Talker adaptation promotes efficient working memory for speech information, especially when listeners must process speech rapidly.
- Talker adaptation potentially reduces cognitive demands during speech encoding and memory retention as reflected in reduced parietal alpha oscillatory power.
- Our results suggest that voice continuity in fast speech allows auditory streaming [5] and reduces cognitive load, whereas non-continuity and slower speech rate may lead to the formation of multiple auditory objects in memory.

This work was supported by grants from the NIH (R03DC014045), Brain and Behavior Research Foundation NARSAD Young Investigator grant to TKP, from the NIH (R01DC009477) to BGSC, and from the NIH (T32DC013017) to SJL.

sungjoo@bu.edu