Talker adaptation or "talker" adaptation? Musical instrument variability impedes pitch perception

Anya E. Shorey [a], Caleb J. King [a], Rachel M. Theodore [b], Christian E. Stilp [a]

[a] Department of Psychological and Brain Sciences, University of Louisville

[b] Department of Speech, Language, and Hearing Sciences, University of Connecticut

Introduction

When hearing the same talker speak, we adapt to that talker's speech: **talker adaptation** (or, talker normalization)

- Listening to speech from a single talker is easier (faster and/or more accurate responses) than multiple talkers

There are several theoretical approaches to talker adaptation including **Active Control** (Heald & Nusbaum, 2014; Magnuson & Nusbaum, 2007), **Episodic** (Goldinger, 1996; 1998), **Bayesian-belief updating** (Kleinschmidt & Jaeger, 2015), **Streaming** (Choi & Perrachione, 2019), and **Efficient Coding** (Stilp & Theodore, 2020)

- All accounts converge on the same prediction: speech judgments are computationally and/or perceptually easier when listening to 1 talker compared to 2+ talkers
- However, speech is not the only structured sound in our environment, and these approaches do not generalize beyond speech equally well

To distinguish among these approaches, we tested and evaluated whether:

1. Adaptation to stimulus structure is speech-specific or general to perception (using musical instruments & tones)
2. The core assumptions of these approaches are sufficiently flexible to accommodate such a test in perception of nonspeech sounds

Hypothesis: Like speech studies, we predicted participants would be faster to respond in the single instrument block than the mixed instrument block, but equally accurate

Method

Participants

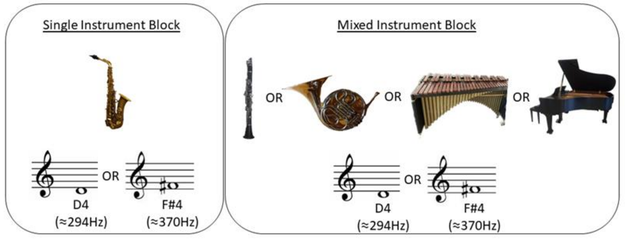
40 undergrad participants

- 36 female, 3 male, 1 other
- *M* = 20.23 years old (95% CI [18.86, 21.69])

Procedure

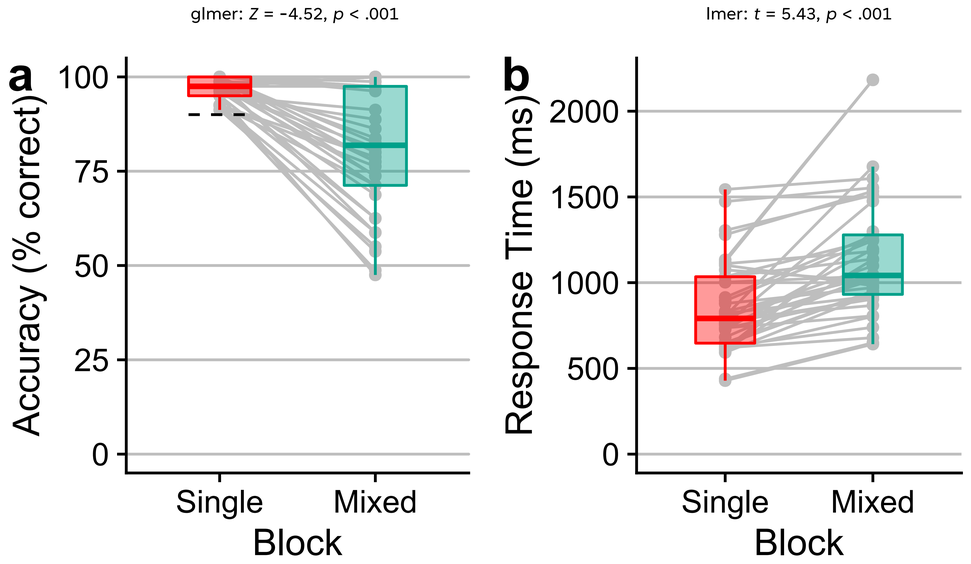
15-minute online study

- Headphone Screen (Woods et al., 2017)
- Exposure to "low" (D4, 294 Hz) and "high" (F#4, 370 Hz) tones on violin
- Practice block: label each violin tone as "low" or "high"
- Main task



- Demographic & musicianship questionnaire

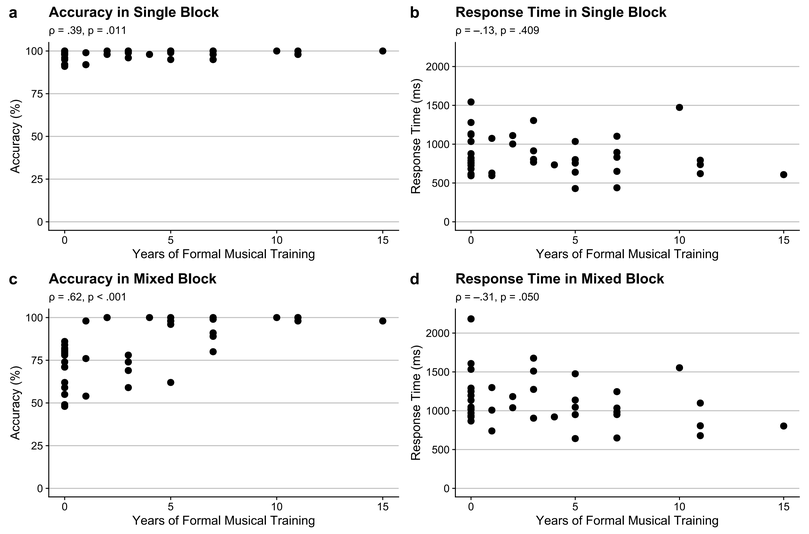
Results



Participants were **faster** and **more accurate** for a single instrument compared to mixed instruments, parallel to speech studies.

Exploratory Analyses of Musical Training

- *M* musical training = 3.53 years (95% CI [2.25, 4.80])



Discussion

Results were consistent with speech studies: adapting to stimulus structure is not limited to talker adaptation

In talker adaptation paradigms, participants are assumed to be expert speech perceivers who exploit existing well-defined categories for speech sounds

- This study doesn't presume that participants have extensive music experience or existing categories for "low pitch" and "high pitch."
- Adaptation might not depend on having long-established categories in place.
- The **streaming** and **efficient coding** approaches do not require such experience, and thus most readily explain benefits from adapting to structure in non-speech sounds.

Limitations

- Task was more difficult than expected for non-musician listeners
  - Perceptual dependence between pitch and timbre (Krumhansl & Iverson, 1994), like that between talker and lexical content (Mullennix & Pisoni, 1990)
  - Accuracy in Mixed Instrument blocks was lower than predicted, but lower accuracy is consistent with the observed longer RTs
- Can't directly consider role of musical training in contributing to these results since participant selection was agnostic toward musical background
- Parallel patterns of results do not definitively indicate identical mechanisms across speech / music perception

Conclusion

- Previous studies reported that structure in speech yields perceptual benefits (i.e., faster and/or more accurate responses to a single talker compared to multiple talkers)
- Some aspects of "talker" adaptation may reflect a general response to structure in the acoustic environment
- Efficient coding and streaming approaches easily accommodate the benefits of structure in less familiar nonspeech stimuli

References

Choi, J. Y., & Perrachione, T. K. (2019). *Cognition*, *192*(June), 103982.

Goldinger, S. D. (1996). *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(5), 1166–1183.

Goldinger, S. D. (1998). *Psychological Review*, *105*(2), 251–279.

Heald, S. L. M., & Nusbaum, H. C. (2014). *Frontiers in Systems Neuroscience*, *8*(MAR), 1–15.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). *Psychological Review*, *122*(2), 148–203.

Krumhansl, C. L., & Iverson, P. (1992). *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 739–751.

Magnuson, J. S., & Nusbaum, H. C. (2007). *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 391–409.

Mullennix, J. W., & Pisoni, D. B. (1990). *Perception & Psychophysics*, *47*(4), 379–390.

Stilp, C. E., & Theodore, R. M. (2020). *Attention, Perception, and Psychophysics*, *82*(5), 2237–2243.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072.