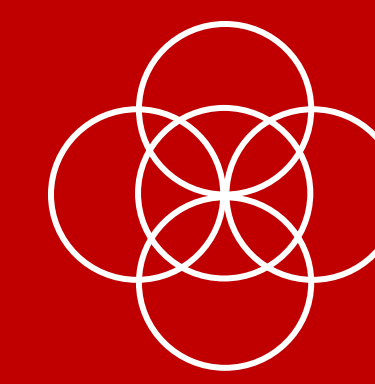


Audiovisual speech, but not talker variability, supports word learning in noise for typically-developing adults

Jasenia Hartman^{1,2}, Jenny Saffran², and Ruth Litovsky²
¹Duke University, ²University of Wisconsin-Madison, USA



WAISMAN CENTER



Symposium of Research in Child Language Disorders, Madison, WI June 2023

Introduction

- Word learning requires listeners to form robust representations of how a word sounds
- However**, most learning environments are noisy, which creates a challenge in perceiving speech sounds¹
- In quiet, talker variability^{2,3} and audiovisual speech^{4,5} help listeners form robust categories of newly learned words

Current Study

Does the co-occurrence of talker variability and audiovisual speech support novel word learning in noise for adults?

Methods

- Participants: 48 young undergraduate adults (18-24 yo)**
 - between and within-subject design

- Looking-while-listening novel word learning task⁶**

3. Learning Conditions

- Talker variability** (between-subject; N = 24/condition)

Single Talker

(assigned speaker counter-balanced across participants)



Multiple Talkers



- Presentation mode** (within-subject; N = 48/condition)

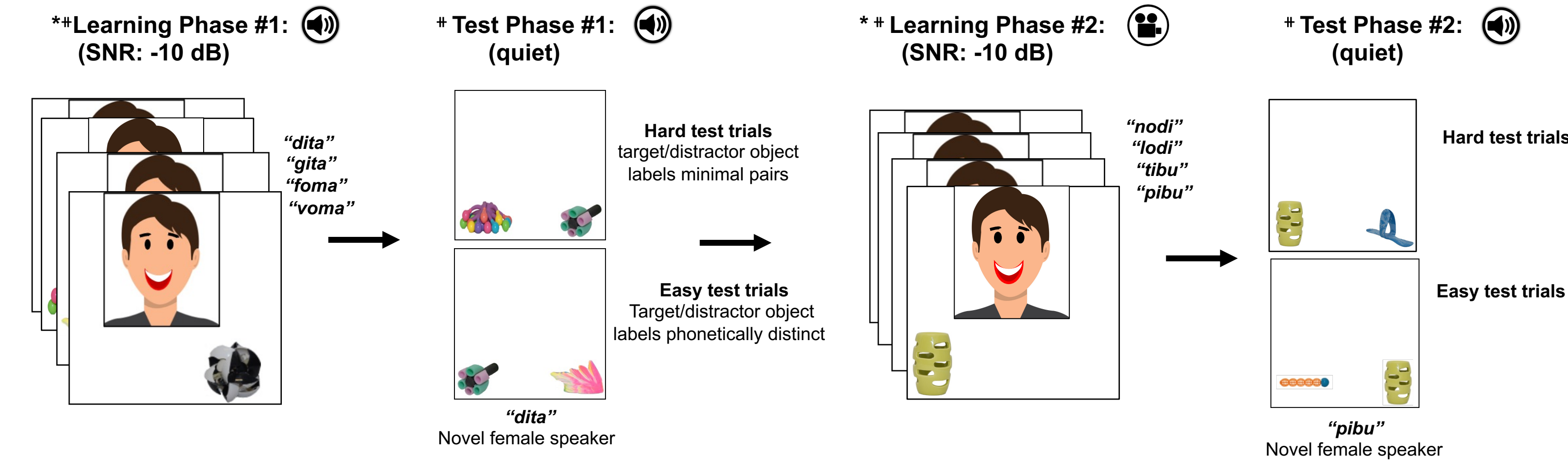
Audio only (still image of talker + audio)

Audiovisual (dynamic video of talker speaking)

- Novel word object pairs** (word set counterbalanced)

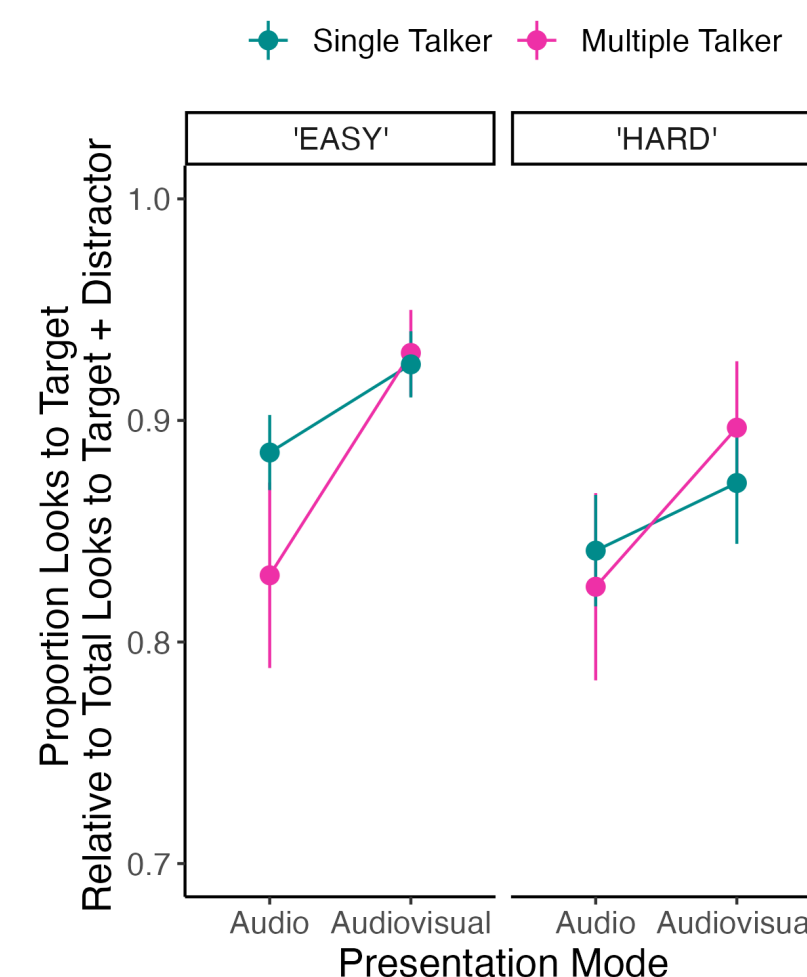
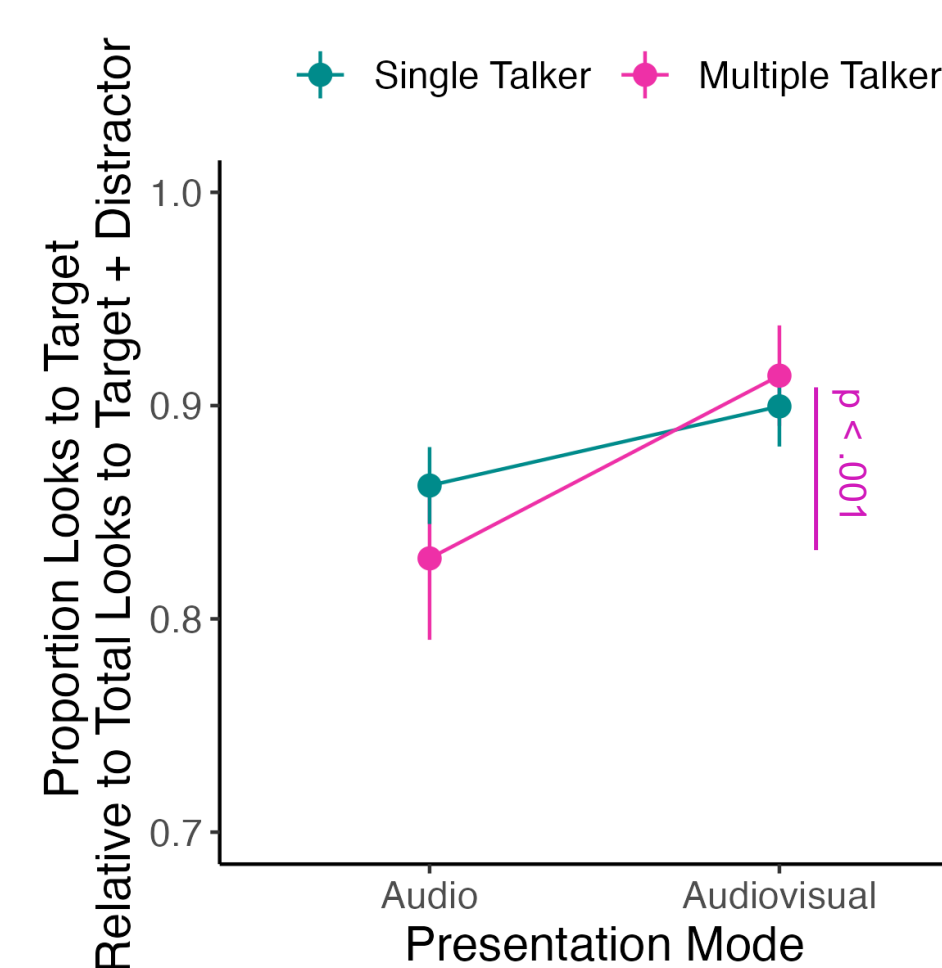
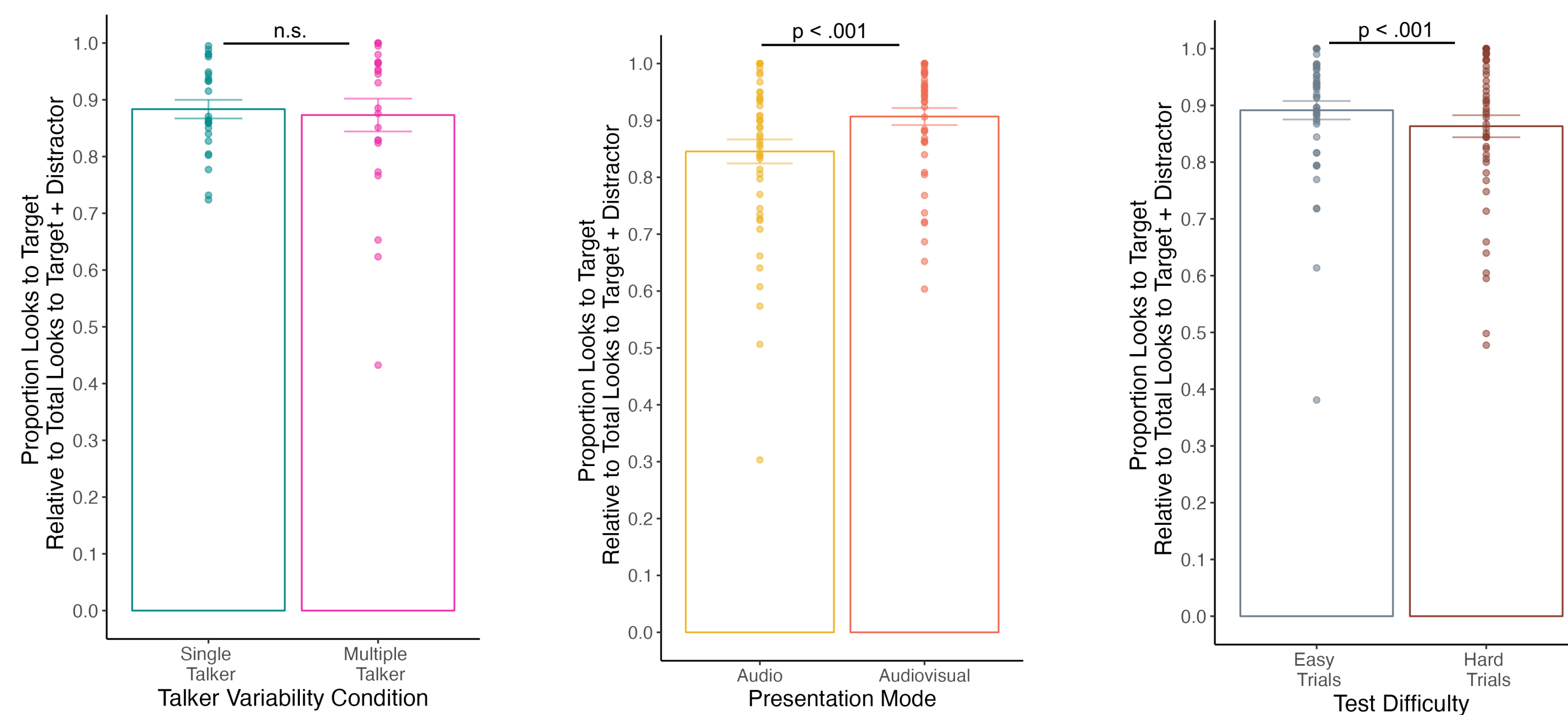
Objects								
Word	/dita/	/gita/	foma/	/voma/	/nodi/	/lodi/	/pibu/	/tibu/
Word Set	Set 1				Set 2			

Procedure (Example of Single Talker Condition)

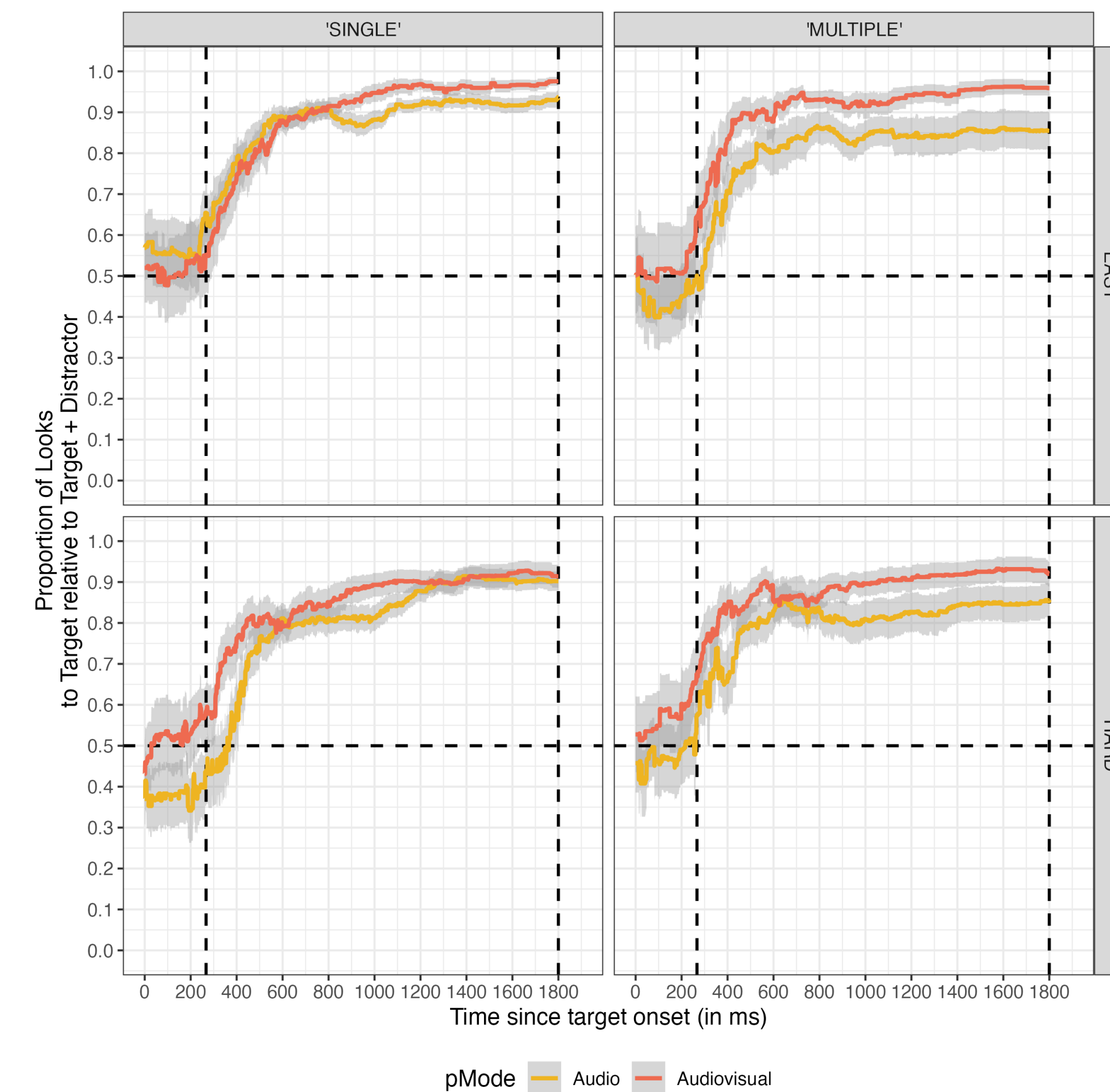


* Presentation mode assignment counterbalanced * 24 total trials (6 trials/target object)

Results: Mean Accuracy



Results: Time Course of Fixations



Conclusions

- Learning from multiple talkers does not boost word learning for adults
 - Adult showed similar performance across talker conditions
- Seeing a talker speak improves word learning in noise more than only hearing the talker, particularly in the presence of acoustic variability
 - Single talker group: performance was similar across presentation modes
 - Multiple talker group: performance was better with audiovisual than audio input

Funding

Diversity Supplement Award (5R01DC016839) in part by a core grant to the Waisman Center (P50HD105353) and NSF GRFP DGE-1256259

Citations:

- Mattys, S.L., Davis, M.H., Bradlow, A.R., & Scott, S.K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27 (7-8), 953-78.
- Rost, G., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12 (2), 339-49.
- Rost, G., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, 15 (6), 608-35.
- Bidelman, G., Sigley, L., Lewis, G.A. (2019). Acoustic noise and vision differentially warp the auditory categorization of speech. *The Journal of the Acoustical Society of America*, 146 (60), 60-70.
- Peelle, J. & Sommers, M (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, 68, 169-81.
- Fernald A, Zangl R, Portillo AL, Marchman V. Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In: Sekerina I, Fernandez E, Clahsen H, editors. *Developmental Psycholinguistics: On-line methods in children's language processing*. Philadelphia: John Benjamins Publishing Company; 2008. p. 97-135

Acknowledgements

We would like to thank Sarah Diel and Lizzy Neubauer for their help with data collection

Contact

Jasenia Hartman: jth68@duke.edu