

Dual-tasking increases reliance on speaker information in phonetic categorization

Helen Reese Klubach and Eva Reinisch, Austrian Academy of Sciences
h.reese@kfs.oeaw.ac.at

Speaker information, such as gender, has long been known to affect the processing of acoustic information, that is, the same sound combined with male and female voices is interpreted differently depending on the gender of the speaker (Strand & Johnson, 1996). In a hierarchy of the order and weighting of information used for speech processing, speaker information could be presumed to be between "low-level" acoustic information and "high-level" lexical information. This is because speaker information is coded in the acoustic signal (rather than words), but speaker identity refers to higher-level knowledge. The addition of cognitive load (i.e., dual-tasking) can be used to test this assumption about levels of processing more directly. Cognitive load (CL) in the form of visual search tasks has been argued to decrease attention to phonetic detail and in turn increase reliance on higher-level information (e.g., lexical bias/Ganong effect; Mattys & Wiget, 2011). However, CL appears not to affect low-level spectral contrast effects (Bosker et al. 2017). The current study examines how the processing of speaker information is affected by the presence of CL and seeks to thereby understand how listeners weigh speaker information. More specifically, we examine if dual-tasking increases listeners' reliance on the gender of the speaker to identify ambiguous sounds on an artificial /s/-/ʃ/ continuum.

A series of experiments was conducted with native speakers of Austrian German. Participants were asked to identify words of two minimal pairs starting with /s/ or /ʃ/ (note that unlike in Germany, Austrian German word-initial /s/ is truly voiceless). A single 15-step continuum was created from /s/ to /ʃ/ which was spliced onto the words produced by one female and one male speaker. Of the 15 steps, nine were presented in the experiments: two end points and seven intermediary steps. Experiment 1 (15 participants) was conducted to replicate Strand and Johnson's (1996) seminal finding that a greater percentage of synthesized stimuli were rated as /s/ when the speaker was male. Experiment 2 (20 participants) tested this effect in combination with CL in the form of a visual search task. Participants had to determine whether a red diamond was present in either a large (8x8 - high CL) or small (3x3 - low CL) grid of red triangles and black diamonds while simultaneously performing the same word identification task as in Experiment 1. They first responded to the auditory stimulus, then to the visual stimulus. The high and low CL conditions were blocked and counterbalanced. Data collection for a third experiment in which participants (20 planned) complete one block with the high-CL task and one block without CL (order counterbalanced) is now underway. The visual grids were shown for 700 ms in all conditions (see Figure 1).

Across experiments we replicated the speaker-effect such that more /s/-responses were given when the speaker was male. Comparing the results between Experiments 1 and 2 also indicates that added CL by dual-tasking significantly increases this effect (see Figure 2). However, no difference was found in the magnitude of the speaker effect between the high and low CL conditions in Experiment 2. This was despite the high-CL task leading to only 69% correct responses and the low-CL leading to 96% correct. Experiment 3 is being conducted to replicate the effect of (high) CL vs. no CL on the processing of speaker gender in a within-participant design.

Overall, our preliminary findings align with the literature assessing the order and weighting of different kinds of information during speech perception. We argued that information about speaker gender could be considered at the boundary of lower-level acoustic and higher-level (e.g., lexical) information. The present set of experiments shows that the processing of speaker information is likely not an entirely low-level, acoustic process. Even if CL impairs attention to low-level phonetic detail, the effect of speaker gender information is not diminished by dual-tasking. Rather, our current results suggest that listeners up-weigh speaker information gathered during speech processing to compensate for decreased attention to acoustic detail.

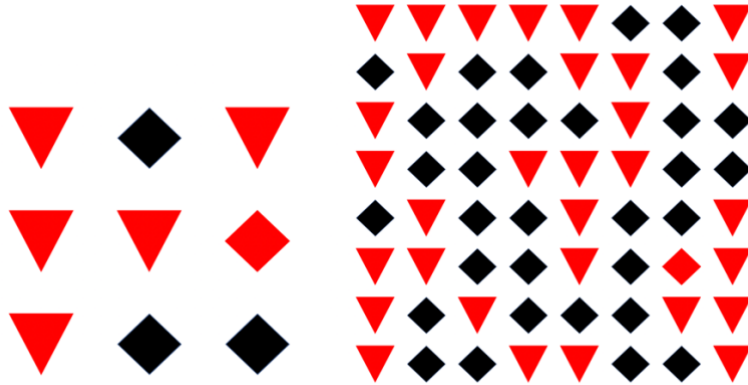


Figure 1. Examples of arrays used for the visual search task. On the left is an example of the 3x3 array used for the low load condition and on the right is an example of the 8x8 array used for the high load condition. Both examples shown here contain an oddball (the red diamond).

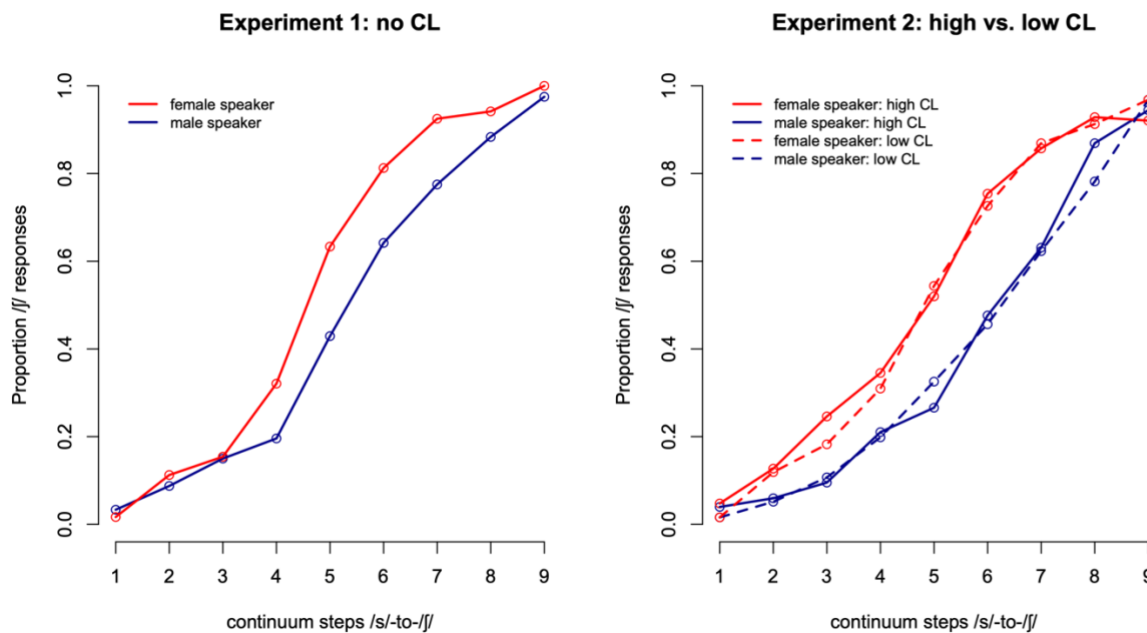


Figure 2. Results from Experiment 1 and 2. On the left, results from Experiment 1 demonstrate that listeners were more likely to classify the same sound as /s/ when combined with a male voice. On the right, this effect is magnified for both the high (solid line) and low (dashed line) cognitive load conditions in Experiment 2.

References

- Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language, 94*, 166-176.
- Mattys, S.L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of memory and Language, 65*(2), 145-160.
- Strand, E. & Johnson, K. (1996). Gradient and Visual Speaker Normalization in the Perception of Fricatives. *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, 14-26.