

SENTENCE-GUIDED RETUNING OF VISUAL PHONETIC CATEGORIES IN AUDIOVISUAL SPEECH

Yun Zou (University of Massachusetts Amherst) & Alexandra Jesse (University of Massachusetts Amherst)
yunzou@umass.edu, ajesse@umass.edu

Speaker idiosyncrasies can affect the perception of phonemes in both auditory and visual speech. To disambiguate and learn about speaker idiosyncrasies, listeners can use lexical knowledge to determine which phoneme the speaker intended to produce and to adjust their phonetic category boundary to recognize the phoneme as intended in future encounters in both modalities (e.g., Norris et al., 2003; van der Zande et al., 2013). However, this use of lexical knowledge seems to fail in the case of idiosyncrasies in word onsets (Jesse & McQueen, 2011). Here, preceding sentence context has been shown to guide the recalibration of *auditory* phonetic categories to adapt to a speaker (Jesse, 2021). In this study, we tested whether sentence context can also retune *visual* phonetic categories. During exposure, forty participants saw and heard a speaker producing sentences. On critical trials, the final word came from an onset /p/-/t/ minimal pair (e.g., *pan* vs. *tan*), but the onset was set to be auditorily and visually ambiguous. For half of the participants, the distal sentence context disambiguated the onset as /p/ (e.g., *Joline enjoys frying eggs for every meal. Unlike others, she never uses inferior cookware for frying.*); for the other half as /t/ (e.g., *Jolene really enjoys the sunshine. Though she never uses any sunscreen, her skin never has a rash.*). The immediate context was the same across conditions within a pair and always supported both interpretations (*Rather, she always has a nice [?]an*). Both groups categorized steps of a visual-only /pa/-/ta/ continuum before and after exposure. The /p/-group gave more /pa/ responses than the /t/-group after exposure, but not before exposure, indicating that the visual phonetic boundary was recalibrated by the distal sentence context during exposure. To summarize, the results of the present study show that sentence context can adjust visual phonetic categories to speaker idiosyncrasies during audiovisual speech perception.

References

- Jesse, A. (2021). Sentence context guides phonetic retuning to speaker idiosyncrasies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(1), 184–194. <https://doi.org/10.1037/xlm0000805>
- Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, 18(5), 943–950. <https://doi.org/10.3758/s13423-011-0129-2>
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. [https://doi.org/10.1016/s0010-0285\(03\)00006-9](https://doi.org/10.1016/s0010-0285(03)00006-9)
- van der Zande, P., Jesse, A., & Cutler, A. (2013). Lexically guided retuning of visual phonetic categories. *The Journal of the Acoustical Society of America*, 134(1), 562–571. <https://doi.org/10.1121/1.4807814>