Coarticulatory Cues in L2 Spoken Word Recognition: A Visual World Eye-Tracking Study

Jéssica Gomes[1], Mercedes Martinez Bruera[2], Kyle Parrish[3], Elise Oltrogge[2]

[1] *CLUL, School of Arts and Humanities, University of Lisbon, Portugal*
[2] *Goethe University Frankfurt, Germany*
[3] *SUNY Brockport, USA*

Background: To understand speech, comprehenders rely on multiple phonetic cues in the acoustic signal as it unfolds. One primary source of phonetic information is coarticulation, a fine phonetic detail in which each sound is influenced by the surrounding sounds [1]. Visual world eye-tracking studies have shown that native speakers make use of coarticulatory features to predict upcoming linguistic elements [2, 3]. However, little is known about whether second language (L2) learners—who have been shown to have fuzzy lexical and phonological representations in their L2 [4]—are similarly sensitive to these fine-grained coarticulatory cues. Through a conceptual replication of [3], we tested whether L2 English learners use coarticulatory cues in spoken word recognition by examining whether the presence of this cue leads to faster target fixations in a visual world eye-tracking task.

Method: We tested 64 German learners of English (mean age of acquisition=7.9 years, SD=1.9; LexTALE mean score=81%, SD=12.7, data collection ongoing) in a visual world eye-tracking experiment. In each trial, participants heard instructions such as "Click on the waffle, please!" while viewing displays containing the target object, an associated distractor, and two unrelated distractors (Fig. 1A). In the no-coarticulatory condition, the determiner and nouns were recorded separately and cross-spliced to remove coarticulatory cues. In the coarticulatory condition, the determiner and noun were produced as a continuous phonological phrase. Following [3], a Gaussian classifier was used to demonstrate that coarticulatory information was present in the determiner's vowel. The eye-tracking data were analyzed using a GAMM-based method [5], which estimated the earliest time point at which looks to the target exceeded looks to the associated distractor (the onset).

Results and discussion: At the group level, the onset for the no-coarticulatory condition was estimated at 225 95% CI [116, 295] ms, and for the coarticulatory condition at 189 95% CI [46, 295] ms (Fig. 1B). The mean difference between the two conditions was 40 ms (95% CI [-96, 204] ms), providing no evidence of a group-level effect. However, at the individual level, LexTale scores significantly predicted fixations to the target image, with participants with higher LexTale scores showing more fixations to the target (Fig. 1C). This effect was stronger in the no-coarticulation condition, suggesting that when cues are absent, and processing demands are therefore higher, proficiency plays a greater compensatory role. Overall, our study contributes to a broader understanding of the extent to which L2 learners encode and use fine phonetic detail during language processing, offering evidence that proficiency may modulate sensitivity to coarticulatory cues in spoken word recognition.

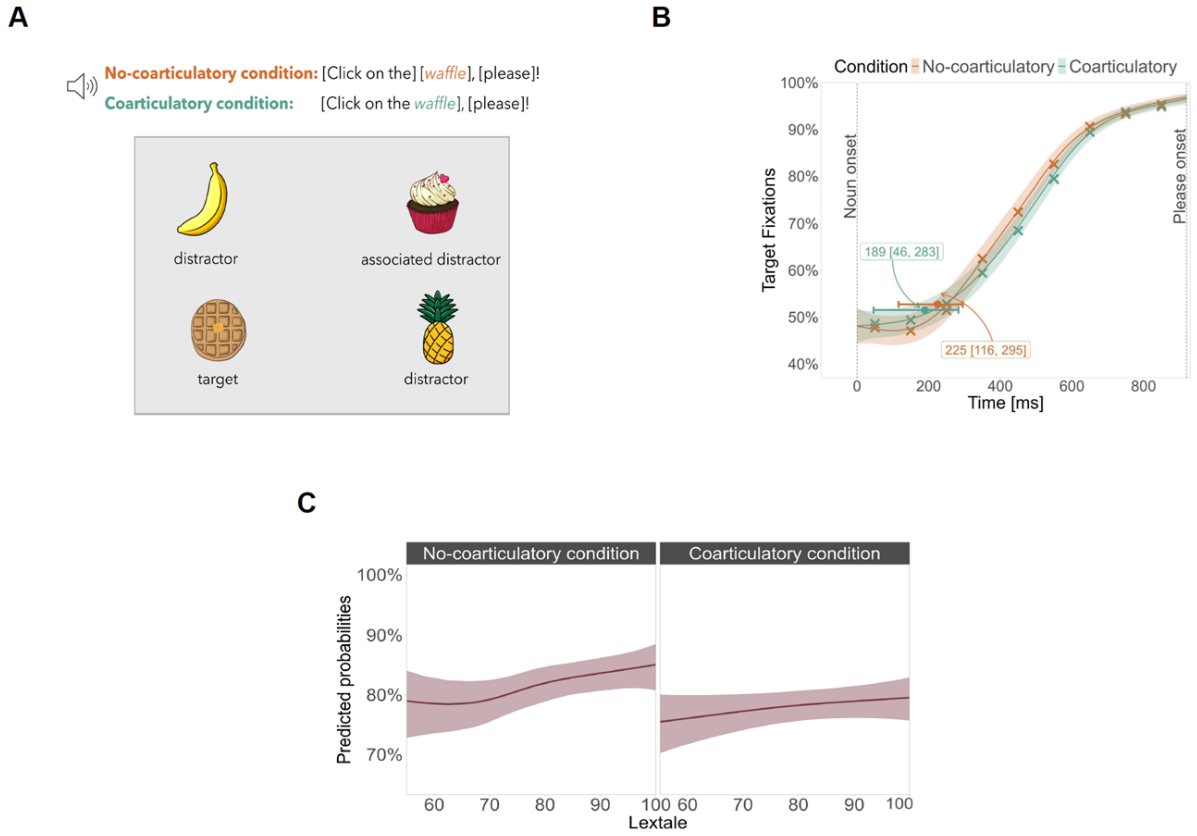

Figure 1: (A) Sample visual display in the experiment. Following [3], one of the three distractors was named "associated distractor" and used to compare with the target object. The target nouns started with an approximant (/r/, /l/, /w/) or with a voiceless fricative (/f/, /s/, /ʃ/). (B) Fitted GAMM curve overlaid onto empirical means of fixation proportions across participants. Points with error bars indicate the median speech-to-eye lag with the 95% highest density interval obtained from the GAMM-based onset detection method. (C) Effect of LexTALE scores on the proportion of fixations to the target image on the no-coarticulatory and coarticulatory conditions.

References

[1] Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*.

[2] Cross, A. M., & Joanisse, M. F. (2018). Eyetracking of coarticulatory cue responses in children and adults. *Language, Cognition and Neuroscience*, *33*(10), 1315–1324. https://doi.org/10.1080/23273798.2018.1484148

[3] Salverda, A., Kleinschmidt, D., & Tanenhaus, M. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/https://doi.org/10.1016/j.jml.2013.11.002

[4] Bordag, D., Gor, K., & Opitz, A. (2022). Refining key concepts of the Ontogenesis Model of the L2 lexical representation. *Bilingualism: Language and Cognition*, *25*(2), 236–241. https://doi.org/10.1017/S1366728921000894

[5] Verissimo, J., & Lago, S. (2026). A novel method for detecting the onset of experimental effects in visual world eye-tracking. https://doi.org/10.31234/osf.io/yk4xb_v4