The interaction between phonetics and phonology when processing the acoustic signal: evidence from labial coarticulation in English and French

Phil J. Howson[1], Marianne Pouplier[1], Francesco Rodriquez[1], Eva Reinisch[2], Justin J. H. Lo[3], Chris Carignan[3], Bronwen Evans[3]

[1]Ludwig-Maximilians-Universität München
[2]Austrian Academy of Sciences
[3]University College London
p.howson@phonetik.uni-muenchen.de

Introduction. English and French both have rounded vowels. However, they differ in the fact that English has phonetic lip rounding on back vowels, while French has phonologically contrastive front rounded vowels. So, the difference between these languages in this regard is that French has phonological specification (Dresher, 2009) of a rounding feature for front vowels. The purpose of this study is to examine how the difference in the hypothesized feature specification affects the perception of anticipatory lip rounding.

In English and French anticipatory lip rounding can be both extensive in nature and extremely variable by speaker (Bell-Berti & Harris, 1982; Vaxelaire, Bonnot, & Keller, 1999; Roy, 2005; Noiray et al., 2010; Howson et al., 2021) and both English and French listeners make in principle use of this information to decode the incoming speech signal (Redford et al., 2018; Sock, Hecker, & Cathiard, 1999; Hirsch et al., 2003). This suggests that English and French language users make use of the incoming phonetic information about lip rounding despite differences in the phonological nature of lip rounding in their languages for front vowels. Linguists have long postulated that phonological contrast in one's language enhances perception (e.g., Trubetzkoy, 1939; Boomershine et al., 2008). Therefore, the presence of phonological contrast for roundedness for front vowels in French but not in English opens up the possibility that anticipatory coarticulation for lip rounding is utilized differently for listeners of each language.

Methods. 22 English and 16 French adult participants were recruited for an eyetracking experiment on the basis that their L1 was either American English or Metropolitan French (hereafter, French). Production data were recorded for all participants. To track the lip shape, we used an adapted version of the "blue lip" technique (Lallouache, 1991). Measurements of lip spread were calculated as the distance between the lip corners (Noiray et al., 2011) and temporal differences were quantified using a sigmoid method (Lo et al., 2023). We chose two degrees of anticipatory coarticulation to measure sensitivity to different amounts of anticipatory lip rounding. Tokens were binned into two groups ("extensive" for relatively longer distance coarticulation and "constrained" for relatively shorter distance coarticulation) based on the distribution of the temporal span of coarticulation in the production data. The stimuli for the perception experiment were chosen from a subset of the production data such that the same speakers contributed to both the extensive and constrained coarticulatory conditions. For English and French, 24 & 29 tokens from the "extensive" (1st quartile) and 24 & 29 tokens from the "constrained" (4th quartile) ends of the distribution were chosen for each language respectively. The mean onset of anticipatory lip rounding preceded the vowel target in English by 402 ms for the extensive category and 136 ms for the constrained category. For French, the extensive category mean preceded the target by 288 ms and by 110 ms for constrained.

During the eyetracking experiment, participants saw a minimal pair of words on the screen which had either an unrounded or rounded target vowel (e.g., English: heed / who'd; French: scie / su) and were instructed to click on the target word as soon as they recognized it. Stimuli were presented in the carrier phrase in which they were produced, since anticipatory coarticulation in the production data was quantified over the entire utterance (English: But Tessa had said *target* pleasantly; French: Mais elle déclarait *target* par hazard). English target stimuli had the vowel pairs /iː/ vs. /uː/, /ɪ/ vs. /ʊ/, /eɪ/ vs/ vs. /oʊ/. French target stimuli had /e/ vs. /ø/, /ɛ/ vs. /œ/, /e/ vs. /o/, /ɛ/ vs. /ɔ/, and /i/ vs. /y/.

Incorrect answers were discarded (< 1% of the data) from analysis and looks to the target or competitor were binned at 5ms intervals. Growth Curve Analysis (Mirman et al. 2008) was used to compute the proportion of fixations on the target. The model included a fixed effect for coarticulation (2 levels: extensive, constrained) and functions for time (time^n, n = [1,7]) The interaction between Coarticulation and Time was also included. Random intercepts were included for participant, speaker, and pair (i.e., the pair of words on the screen). We computed one model for English and one for French. To determine differences in growth curves for extensive and constrained conditions, we computed a smoothing spline for each model by randomly sampling participants from each condition and fitting a smoothing curve to their data. This was done 1000 times to generate a distribution and obtain 95% confidence intervals (Wendt et al., 2014).

Results. The results of the GCA for English revealed that there was an increase in target fixations at approximately 75 ms after the onset of the target segment, but no significant difference for the interaction between coarticulation and functions of time (time^n: $p > 0.05$; Figure 1). Given that eye movement planning and execution lags the input stimuli by approximately 200 ms (Travis, 1936), we added 200 ms to the spike in fixations observed in the results to estimate when word recognition took place. English listeners utilized coarticulation at approximately 125 ms before the target vowel onset. The analysis of French, on the other hand, did reveal a significant difference for the interaction between

coarticulation and functions of time (time^n: $p < 0.05$, except time^5: $p = 0.82$) and revealed an increase in fixations towards the target 50 ms before the onset of the critical target in the extensive condition and 110 ms after the onset of the target stimuli in the constrained condition (Figure 1). French listeners had a rapid increase in looks to the target at roughly 200 ms after the onset of coarticulation in both conditions. When taking saccade lag time into account (~200 ms), this suggests listeners recognized the target at approximately 250 ms (extensive condition) and 90 ms (constrained condition) before target vowel onset. This indicates that French listeners use the coarticulation related to lip rounding as soon as it is available in the speech stream, whether that is extensive or more constrained anticipatory coarticulation.

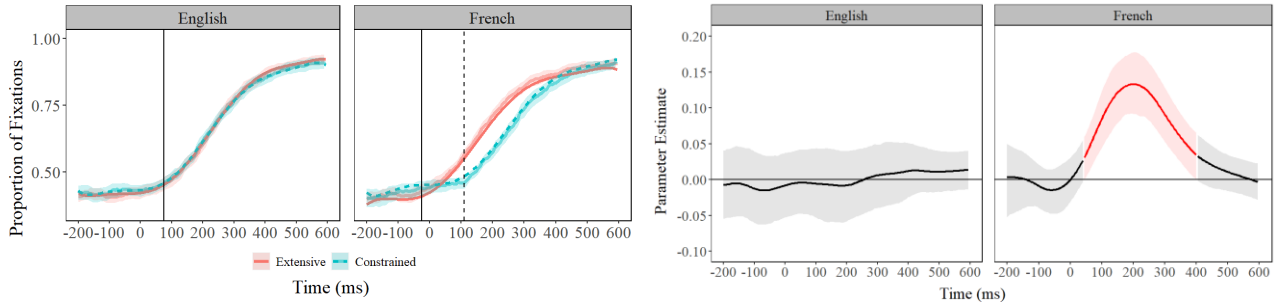

Figure 1 (left): Growth curve analysis for extensive (red) and constrained (blue) for English (left) and French (right). The mean Proportion of Fixations for extensive (opaque red) and constrained (opaque blue) with ±1 SE are also presented. Vertical lines indicate estimated point of increase in fixations on the target for extensive (solid) and constrained (dashed) coarticulation. Figure 1 (right): divergence plots comparing the difference between extensive and constrained for English (left) and French (right). Red indicates, in the right-hand graph, a significant difference between contours. 0 ms indicates the onset of the critical segment ((un)rounded vowel).

Discussion. The results indicated English listeners start to recognize anticipatory lip rounding at around 125 ms before the target vowel onset. The French listeners on the other hand recognize upcoming lip rounding as far back as 250 ms before the target onsets and displayed sensitivity to differences in extensive and constrained coarticulation. The difference in perception contrasts with the differences in the stimuli: English had a mean coarticulatory onset of 402 ms (extensive) and 136 ms (constrained) before the target vowel, while French had a mean onset of 288 ms (extensive) and 110 ms (constrained) before the target vowel. So, despite the availability of coarticulatory information earlier in the speech stream for English than in French, English listeners did not show any differences between their perception of extensive and constrained coarticulation. Additionally, English listeners did not demonstrate an increase in fixations as early as French listeners. The reason for this is possibly due to the phonological status of lip rounding in French. The data thus supports the notion that the presence of a phonological contrast improves sensitivity to subtle differences in coarticulation related to the acoustic-phonetic cues to that contrast. Whether French listeners are equally sensitive to rounding in front-back vowel pairs will have to be addressed in future research.

References

Bell-Berti, F. & Harris, K. S. (1982). Temporal patterns of coarticulation: lip rounding. *Journal of the Acoustical Society of America*, 449-454.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In Avery P., Dresher E., & Rice K. (eds.), *Contrast in Phonology*, pp. 143-172. Berlin: Mouton de Gruyter.

Dresher, B. E. (2009). The Contrastive Hierarchy in Phonology. Cambridge University Press, Cambridge.

Hirsch, F., Sock, R., Connan, P., & Brock, G. (2003). Auditory effects of anticipatory rounding in relation with vowel height in French. *Proceedings of the International Congress of Phonetic Sciences*, 1445-1448.

Howson, P. J., Kallay, J. E., & Redford, M. A. (2020). A psycholinguistic method for measuring coarticulation in child and adult speech. *Behavior Research Methods*, 846-863.

Lallouache, M. T. (1991). *Un poste visage-parole couleur: Acquisition et traitement automatique des contours des lèvres*. Ph.D. dissertation, Institut National Polytechnique de Grenoble.

Lo, J. J.H., Carignan, C., Pouplier, M., Alderton, R., Rodriquez, F., Evans, B. G., & Reinisch, E. (2023). Language specificity vs speaker variability of anticipatory labial coarticulation in German and English. *Proceedings of the International Congress of Phonetic Sciences*, 2105-2109.

Ménard, L., Cathiard, M., Troille, E., & Girouxx, M. (2016). Effects of congenital visual deprivation on the auditory perception of anticipatory labial coarticulation. *Folio Phoniatrica et Logopaedica*, 83-89.

Noiray, A., Cathiard, M., Abry, C., & Ménard, L. (2010). Lip rounding anticipatory control: Crosslinguistically lawful and ontogenetically attuned. In Maassen, B. & van Lieshout, P. (eds.), *Speech motor control: New developments in basic and applied research*, pp. 153-171. Oxford, United Kingdom: Oxford University Press.

Redford, M. A., Kallay, J. E., Bogdanov, S. V., & Vatikiotis-Bateson, E. (2018). Leveraging audiovisual speech perception to measure anticipatory coarticulation. *Journal of the Acoustical Society of America*, 2447-2461.

Roy, J. (2005). Visual perception of anticipatory rounding gestures in French. *Proceedings of Interspeech*, 2949-2952.

Sock, R., Hecker, V., & Cathiard, M. (1999). The perceptual effects of anticipatory labial activity in French. *Proceedings of the International Congress of Phonetic Sciences*, 2057-2060.

Travis, R. C. (1936). The latency and velocity of the eye in saccadic movements. *Psychological Monographs*, 242-249.

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie* [= TCLP 9]. Československo: Praha.

Vaxelaire, B., Sock, R., Bonnot, J., & Keller, D. (1999). Anticipatory labial activity in the production of French rounded vowels. *Proceedings of the International Congress of Phonetic Sciences*, 53-56.

Wendt, D., Brand, T., & Kollmeier, B. (2014). An Eye-Tracking Paradigm for Analyzing the Processing Time of Sentences with Different Linguistic Complexities. *PLoS ONE*, e100186.