Comparing the real-time perception of French nasal and labial coarticulation

Francesco Rodriquez[1], Marianne Pouplier[1], Phil J. Howson[1], Eva Reinisch[2], Justin J.H. Lo[3], Christopher Carignan[4], Bronwen G. Evans[4]

[1]Ludwig-Maximilians-Universität München, Germany
[2]Austrian Academy of Sciences, Vienna, Austria
[3]Lancaster University, United Kingdom
[4]University College London, United Kingdom

{f.rodriquez|pouplier|p.howson}@phonetik.uni-muenchen.de,
Eva.Reinisch@oeaw.ac.at, j.h.lo@lancaster.ac.uk,
{c.carignan|bronwen.evans}@ucl.ac.uk

Introduction. An integral feature of speech production is the overlap of articulatory gestures in time, a phenomenon also known as coarticulation. Coarticulatory information permeates the acoustic signal and is actively sought out by listeners in spoken-word recognition (Fowler & Brown 2000). In French, it is reported that fine-grained differences in coarticulatory timing for both nasality and lip rounding provide perceivable information to listeners for phoneme identification and disambiguation (Desmeules-Trudel & Zamuner 2019; Benguerel & Adelman 1976; Hirsch et al. 2003). An intuitive explanation would be to attribute the sensitivity to these coarticulatory cues to the contrastive nature of nasality and lip rounding in the French vowel system. However, anticipatory lip rounding and nasal coarticulation are known to differ in their temporal extent in French: while anticipatory lip rounding was found to be quite extensive and variable (e.g., Noiray et al. 2010) the temporal extent of nasal coarticulation is rather constrained (e.g., Pouplier et al. 2023). This difference in the temporal extent of anticipatory coarticulation within the same language affords the unique opportunity to investigate whether these inter-articulator production differences are directly reflected in listeners' perceptual patterns. Also, the question of whether information from coarticulation with the same phonological status (here: contrastive) can be expected to be treated in the same way by listeners has not been directly investigated yet. The results of such an inquiry could shed light on the complex dynamics of the production-perception link in coarticulation (e.g., Beddor et al. 2018).

The present study investigates the use of coarticulatory timing cues in French by examining listeners' perceptual sensitivity to temporal variation in anticipatory nasalization and anticipatory lip rounding. We ask whether listeners are sensitive to fine-grained temporal differences in coarticulatory timing during spoken-word recognition as soon as the information becomes available in the signal. Listeners' responses should then systematically vary with coarticulatory onset for both articulators.

Methods. 16 native speakers/listeners of French participated in both a production (reading task) and a perception experiment (visual-world eye-tracking experiment). *Production experiment:* Nasal coarticulation data was recorded with a nasalance device and nasality was measured as (mean-normalized) nasal channel intensity. Anticipatory lip rounding data was quantified off video recordings (Lallouache, 1991). Rounding was measured as the distance between the lip corners (lip spread). The target stimuli, embedded in a carrier phrase, contain minimal pairs distinguished by a nasal/oral consonant in VN/VC sequences (e.g., *l'aîné* [le.ne] vs. *l'été* [le.te]) or by a rounded/unrounded vowel (e.g., *Caire* [kɛʁ] vs. *cœur* [kœʁ]) of the following front vowel pairs: /e/~/ø/, /ɛ/~/œ/, /i/~/y/. The onset of coarticulation was determined algorithmically (as explained in Lo et al. 2023) as a point in signal divergence between a given nasal/oral (rounded/unrounded) minimal pair.

Perception experiment: Tokens for the eye-tracking experiment were selected per articulator by sampling the extremes ($1^{st}/4^{th}$ quartile) of the coarticulation onset point distributions across speakers. For each articulator the same number of tokens with extensive and constrained coarticulation was selected. The extensive category's mean coarticulation onset preceded that target segment by 288ms for lip rounding and by 154ms for nasality. The constrained category's mean coarticulation onset preceded the target segment by 110ms for lip rounding and 47ms for nasality. Each target was presented in its original carrier phrase (114 unique trials), each presented twice and mixed with about the same number of fillers. Visual referents were printed words of the minimal pairs with targets presented once on the left and once on the right. Participants clicked on the word that they heard (two-alternative forced choice task). The probability of target fixations over time (5ms bins) was calculated by means of Growth Curve Analysis (GCA) (Mirman et al. 2008). One model per articulator was computed including a fixed effect of coarticulation type (extensive, constrained), functions for time ($time^n$, $n = [1,7]$) as well as the interactions between the two. Smoothing splines were used to obtain 0.95 confidence intervals for fixation proportions over time as well as to determine differences in the growth curves for the extensive and constrained condition (Wendt et al. 2014).

Results. The growth curves of the proportions of fixation on the correct target word are shown in **Figure 1** (A-B). The extensive and constrained conditions for an articulator are compared in divergence plots (**Figure 1**, C-D), where significant divergences in correct target fixations between the extensive and constrained condition are marked in red. For nasality, the target fixation curve in the extensive condition rises ~50ms after target segment onset. Considering 200ms needed to perform a saccade (Travis 1936), the results suggest that listeners use extensive nasalization cues as soon as they are available in the signal (mean anticipatory interval of nasality in extensive condition: 154ms) (**Figure 1**, A). Responses in the constrained condition align with the extensive condition, so the curves do not significantly diverge (**Figure 1**, C). This is surprising as it greatly exceeds the time window at which nasalization cues become available in the constrained condition (anticipatory interval: 18ms to 59ms). Possibly, this is due to confounds or unresolved biases currently under investigation. For lip rounding, target fixations in the extensive condition start rising at target onset (= zero point), suggesting a listener response slightly delayed with respect to the mean onset of coarticulation (mean anticipatory interval of rounding in extensive condition: 288ms) (**Figure 1**, B). The constrained condition significantly diverges from the extensive condition at 50ms into the target (**Figure 1**, D).

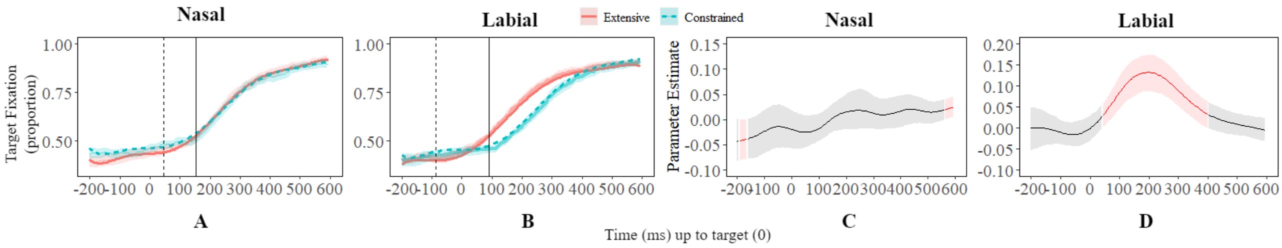

Figure 1: *Growth curve analysis for extensive (red solid line) and constrained (blue dotted line) tokens for nasal (A) and labial (B) perception. Mean fixation proportions in opaque colors. Estimated earliest point of reaction in perception (mean coarticulatory onset up to target across produced stimuli + 200ms for saccade execution), for the extensive (vertical dashed line, A-B) and constrained condition (vertical solid line, A-B). Divergence in proportion of fixation between extensive and constrained coarticulation for nasal (C) and labial (D) coarticulation. Significant differences are marked in red. Target onset at 0.*

Discussion. The results suggest that French listeners are sensitive to extensive coarticulatory cues as soon as they become available in the acoustic signal for both nasal and labial coarticulation. The time course of perception thus mirrors the fine-grained differences in the temporal extent of coarticulatory information in production. This observation might support the idea that French listeners are attuned to anticipatory nasalization and lip rounding as real-time cues because of their phonologically contrastive status. Notably, in French this seems to hold despite the observation that anticipatory lip rounding is more variable (i.e., the cue is less consistent) than anticipatory nasalization. While we focused on group-level perceptual patterns here, in planned analyses we intend to compare speaker-specific production-perception patterns (as in Beddor et al. 2018) to test whether an individual speakers' coarticulatory behavior in production (e.g., early vs. late 'coarticulators') is reflective of their use of coarticulatory information in perception (e.g., early vs. late use of coarticulatory cues).

References

Beddor, P., Coetzee, A., Styler, W., McGowan, K., & Boland, J. (2018) The time course of individuals' perception of coarticulatory information is linked to their production: implications for sound change. Language, 94(4), 931-968.

Benguerel, A.P., & Adelman, S. (1976). Perception of coarticulated lip rounding. Phonetica, 33(2), 113-126.

Desmeules-Trudel, F., & Zamuner, T.S. (2019). Gradient and categorical patterns of spoken-word recognition and processing of phonetic details. Attention, Perception, & Psychophysics, 81(5), 1654-1672.

Fowler, C.A., & Brown, J.M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. Perception & Psychophysics, 62, 21-32.

Hirsch, F., Sock, R., Connan, P., & Brock, G. (2003). Auditory effects of anticipatory rounding in relation with vowel height in French. Proceedings of the 15th International Congress of Phonetic Sciences, 1445-1448.

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 20th International Congress of Phonetic Sciences,* 2105-2109.

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

Mirman, D. Dixon, J.A., & Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: growth curves and individual differences. *Journal of Memory and Language, 59*(4), 475-494.

Pouplier, M., Rodriquez, F., Alderton, R., Lo, J.J.H., Reinisch, E., Evans, B.G., Carignan, C. (2023). The window of opportunity: Anticipatory nasal coarticulation in three languages. *Proceedings of the 20th International Congress of Phonetic Sciences* (Prague, Czech Republic), 2085-2089.

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*, Oxford, United Kingdom: Oxford University Press, 153-171.

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