Anticipatory coarticulation: a psycholinguistic perspective and method

Melissa A. Redford[1], Sergei V. Bogdanov[2], Eric Vatikiotis-Bateson[3]
[1]University of Oregon, [2]Moonshadow Mobile, [3]University of British Columbia

The speech plan is an abstract control structure that encodes units of production and their phonetic form in sufficient detail to drive speech action. Long-distance anticipatory coarticulation has long provided evidence for such a control structure (e.g., Kozhevnikov & Chistovich, 1965), but the phenomenon has rarely been investigated in such a way as to distinguish between different theories of production, which nonetheless abound. To take just one example, a class of derivational production models assume production units based on metrical structure (Levelt et al. 1999; Keating & Shattuck-Hufnagel, 2002; Shattuck-Hufnagel, 2015); an alternative view is that production units larger than the word coincide with syntactic structure (Selkirk, 1996; Redford, 2015); other models assume that anticipatory coarticulation results from articulator-specific timing patterns, which are independent of prosodic or syntactic boundaries (e.g., Bell-Berti & Harris, 1981; Byrd & Saltzman, 2003; Goldstein et al., 2007). Accurate temporal measures of anticipatory coarticulation and careful control over speech stimuli provide a means for distinguishing between the various theories. The present study was undertaken to demonstrate this and to validate a non-invasive psycholinguistic method for studying anticipatory coarticulation.

Our method builds on work in AV speech perception (Munhall & Tohkura, 1998; Moradi et al., 2013). We use a gating paradigm to assess the temporal onset of anticipatory lip rounding across two metrical contexts. Five adult speakers were recorded while producing multiple repetitions of different minimal pair SVO sentences where the object consisted of a determiner, "the," and monosyllabic noun with a rounded or unrounded rhyme ("oop" or "ack" as in *soup* and *sack*). Metrical context was varied by using a phrasal verb ("packs up") or a plain verb ("packs"). The subject was held constant ("Maddy"). Once recorded, sentences were isolated and gated based on acoustic landmarks from the midpoint of the verb through to the midpoint of the object noun. Judges (10 perceivers per speaker) were asked to listen to full and partial versions of the sentences produced by a speaker and decide whether the object noun rhymed with rounded "oop" or unrounded "ack." If production units are determined by metrical structure, then anticipatory lip rounding should be evident at the onset of "the" in sentences with phrasal verbs, but not in sentences with plain verbs. If production units align with syntactic structure, then anticipatory lip rounding should align with the onset of "the" regardless of metrical context. If gestural activation is insensitive to prosodic or syntactic boundaries, then the onset of lip rounding for "oop" should depend on phonetic context and be evident earlier in the plain verb case than in the phrasal verb case, due to the biliabial offset in the phrasal verb. The results indicated significant effects of sentence type (phrasal verb vs. plain verb) and test gate on anticipatory lip rounding, but no interaction between sentence type and test gate on the critical gates associated with "the" onset and nuclei (see Figure 1). The results are therefore most consistent with the view that production boundaries align with syntactic boundaries. Results also indicated earlier correct identification of rounding in the AV condition compared to a control, audio-only condition (also 10 perceivers per speaker), suggesting that AV judgments provide a more sensitive measure of anticipatory coarticulation than could be obtained from examining speech acoustics alone.

The AV gating method allows us to test specific predictions regarding the onset of lip rounding by defining different gates based on specific temporal or segmental (= acoustic) criteria. Though somewhat time- and resource-consuming, the method allows us to take sensitive measures of articulatory movement without attaching sensors to a speaker's face or speech articulators. Moreover, in contrast to algorithmic analyses of speech movement based on video, the AV method is robust to co-speech head movements (see Munhall et al., 2004). This feature in particular makes the method valuable for testing specific hypotheses regarding the structure of the speech plan in different populations of speakers, which is how we intend to use it.

References

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149-180.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, 103, 386-412.

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 101, 112-156.

Kozhevnikov, V. A.,& Chistovich, L.A. (1965). *Speech: articulation and perception* (revised). (Joint Publications Research Service, 30,543.) Washington: U.S. Department of Commerce.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-38.

Moradi, S., Lidestam, B., & Rönnberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: effects on time and accuracy. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2013.00359

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility head movement improves auditory speech perception. *Psychological Science*, 15, 133-137.

Munhall, K. G., & Tohkura, Y. (1998). Audiovisual gating and the time course of speech perception. *Journal of the Acoustical Society of America*, 104, 530-539.

Redford, M.A. (2015). Unifying speech and language in a developmentally sensitive model of production. *Journal of Phonetics*, 53, 141-152.

Shattuck-Hufnagel, S. (2015). Prosodic frames in speech production. In M.A. Redford (Ed.), *The handbook of speech production* (pp. 419-444). Boston, MA: Wiley.

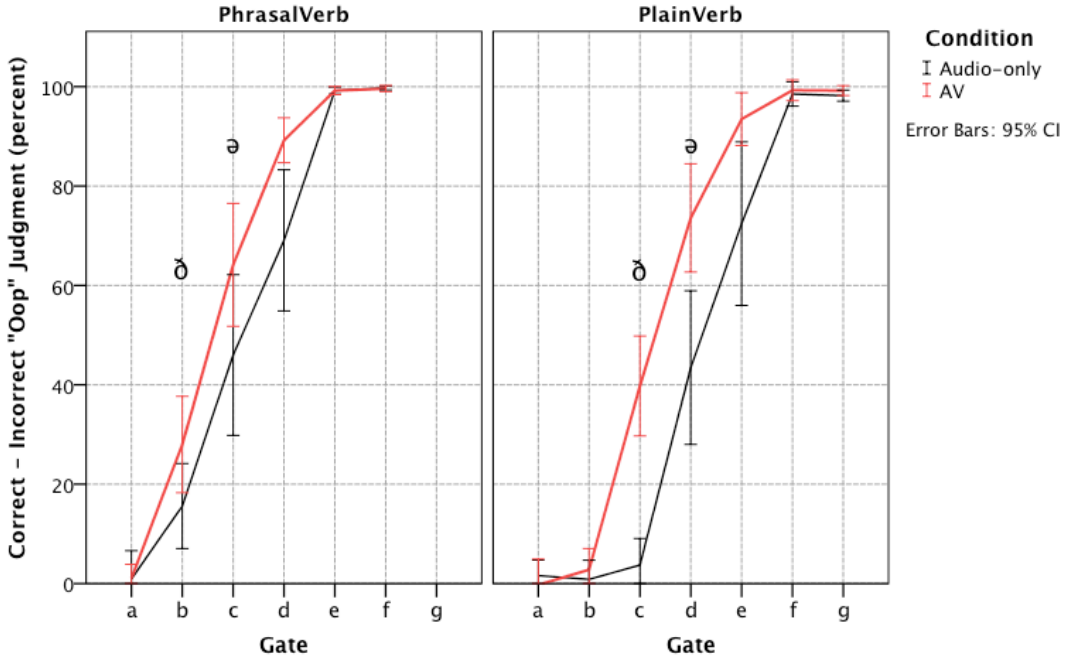

Figure 1. Bias-corrected identification of anticipatory lip rounding in AV or audio-only speech at specific temporal locations (gates) across two metrical contexts (phrasal verb vs. plain verb).