REVISITING CV TIMING WITH A NEW TECHNIQUE TO IDENTIFY INTER-GESTURAL PROPORTIONAL TIMING

Karthik Durvasula and Yichen Wang

Michigan State University
durvasul@msu.edu, wangy176@msu.edu

ABSTRACT

In this article, we probe the timing relationship in CV sequences. To do so, we develop a new technique that is sensitive to the distinction between proportional lag relationships and constant lag relationships, which is a problem for previous explorations of CV lag. Using the new technique, we show that our data is consistent with the vowel onset being synchronised with the preceding consonant offset (what we call "offset-onset alignment"). We also observed more stability with target onset alignment than with gesture onset alignment. Our results suggest that speakers might be planning target achievement more so than alignment of gesture onsets. Furthermore, they suggest a need for a new explanation for the C-centre effect.

Keywords: CV timing, in-phase, anti-phase, offset-onset alignment, X-Ray microbeam data.

1. INTRODUCTION

In Articulatory Phonology [1, 2, 3, amongst others], inter-gestural timing between two gestures (G_1 and G_2) boils down to one of three phasal relationships, where an internal clock cycle of a gesture is represented by 0°-360°: (a) *in-phase*: when gestures start simultaneously (0°phase difference); (b) *anti-phase*: when one gestural onset is perfectly asynchronous with another gestural onset (180° phase difference); (c) *eccentric phase*: gestures have a phasal relation that is different from 0° and 180°.

With the above timing relationships as backdrop, one can layout three different views of CV timing that have been directly or indirectly argued for in prior research. The first hypothesis is what might be called the standard position [1, 2, 4] — the consonant gesture (G_1) and the vocalic gesture (G_2) are in-phase (Figure 1a). When an additional claim that consonants in CCV sequences are in an anti-phase relationship is added, then this first viewpoint of CV timing is able to predict that C-CENTRE-TO-ANCHOR interval stability pattern wherein the consonantal gestures in a complex onset are in a global timing relationship with the subsequent vowel [5, 6, 7, 8, amongst others].

A second hypothesis about CV timing depends on a generalisation of the *split-gesture hypothesis*, whereby a stop gesture is decomposed into two gestures related to the closure and release portions, respectively [9]. This hypothesis was used by Nam [9] to account for the fact that impressionistic observations of prior work suggested that the *centre* of the observed stop gesture, and not the onset, was in a stable temporal relationship with the following vowel. A similar view of a separately manipulable neutral attractor gesture for all segment gestures was discussed in earlier research work [10, 11, 12]), and, in fact, a similar observation about the timing of CV sequences that, impressionistically, the V gesture starts roughly mid-way through the observed consonant gesture can be seen in some of the earliest work on gestural timing [5] not just for stops, but also liquids and fricatives. Therefore, one could perhaps generalise Nam's split-gesture hypothesis to all segments, and claim that vowel gestures start mid-way through any preceding consonantal gesture. However, the actual simulations that Nam presented do not bear this prediction out, even for stops. His simulations predict that the vowel gesture has a 60° phase difference with a preceding consonant (stop) closure gesture and a phase difference of -60° with a preceding (stop) release gesture, when an equal coupling strength is assumed for all timing relationships. Consequently, it is straightforward to show that, if each gesture needs about 240°-360° of its internal clock duration to attain the target (see [2] for the lower bound), the vowel itself is expected to start about 12.5%-16.7% of the entire observed consonant (or specifically, stop) gesture duration, and not mid-way through the consonant gesture (derivation not presented in the interest of space).

A third view about CV timing can be seen as a generalisation of the work by Shaw, Durvasula, Kochetov and Oh [13, 14] who argue that, in consonant sequences, G_2 is aligned to the end of G_1 — we call this an *offset-onset alignment*

relationship. If we were to generalise this relationship to all segment sequences, we would predict that in CV sequences, the vowel gesture is in an *offset-onset alignment* relationship with the preceding consonant gesture, *i.e.*, the proportional lag is 100% between the two gestures (Figure 1b).

Most of the evidence furnished in support of in-phase timing in CV sequences is either based on observed lag (see next section) or based on general work on motor control [15, 10]. For example, in a detailed quantitative/experimental evaluation of this question done quite recently, [16, 4] develop a new experimental paradigm that uses triplets of stimuli (*e.g.*, [lolju] vs. [loju] vs. [lolu], which allowed them identify the actual gestural onset of the consonant and vowel gestures through a comparison of minimal pairs. They found that the labial gesture of an [u] starts in-phase with the preceding consonant [l] gesture, but the tongue tip gesture of the same vowel starts towards the end of the consonant gesture. They interpret this result as consistent with in-phase CV timing, with the caveat that a vocalic gesture (*e.g.*, tongue tip) that conflicts with the preceding consonant gesture is delayed and starts during the offset of the consonantal gesture. The comparative technique they employ has long been argued to allow one to separate out the effect of context [17, 18, 19]; however, there are three issues with their technique/interpretation. First, based on their data, one could have also argued for the opposite conclusion that the vowel gesture starts during the offset of the consonantal gesture, but the rounding of the vowel gesture starts earlier, *i.e.*, the in-phaseness is specific to rounding. In fact, Benguerel and Cowan [17] observe that rounding in French starts up to 6 segments before the vowel (across syllables and even words), even when one uses a comparative technique to identify the onset of the gesture. So, it is possible that the rounding pattern is simply a language-specific fact or assimilation in English and French. Second, an issue with the comparative technique is that it requires *phonetic* minimal pairs, and not phonological minimal pairs. For example, if there is fronting of [u] after [l] (see [20]), then [lu] doesn't form a good minimal pair with [ju]. Third, the technique doesn't allow us to separate absolute lag from proportional lag (see next section).

2. A NEW WAY TO IDENTIFY PROPORTION LAG BETWEEN TWO GESTURES

The main technical issue that we address in this paper is that any observed lag (*i.e.*, the actual duration of time) between gestures can be due to two different sources: (a) a proportional (relative) temporal lag between the two gestures, wherein G_2 starts at a certain proportional point of G_1. Note, this is the temporal relationship represented by the phasal relationships discussed above; (b) an absolute (or constant) temporal lag between two gestures, wherein G_2 starts at a constant time in relation to G_1. Absolute lag may be a result of planning, or biomechanical factors [21, 8], or due to differences in measurement errors related to different gestures.

Given that there are two sources that can contribute to the observed lag between gestures, one can't simply look at the observed average lag between two gestures to establish the phasal relationship between them. For example, G_2 onset could be timed in-phase with G_1 but with a positive absolute lag (Figure 1c), or G_2 onset could be timed to the end of G_1 but with a negative absolute lag (Figure 1d). Both the alternatives would superficially look the same. If we calculated the average observed lag between the onsets of the two gestures and then used that duration to calculate the proportional lag (compared to G_1 duration), we would get the same proportional lag value in both cases. This is quite problematic as the cases have quite different underlying temporal organization of the gestures, while one has in-phase alignment the other has offset-onset alignment.

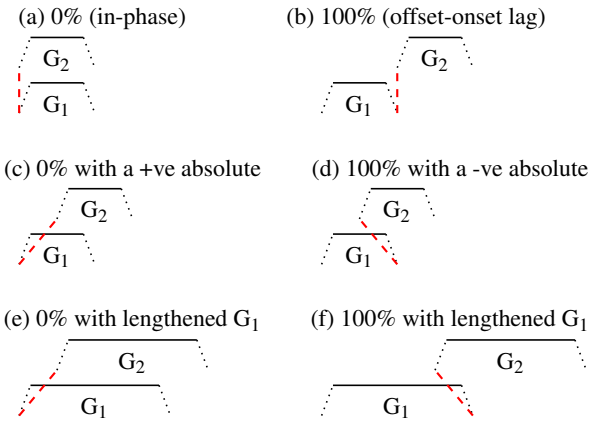

Figure 1: Different timing possibilities between two gestures and their consequences for a lengthened G_1 (proportional lag point = red dashed line)

One way to solve this measurement issue is to recognise that the two possibilities in Figure 1c-d make different predictions for how the simple lag duration *co-varies* with a change in G_1 duration. In the first case (represented in Figure 1e), the observed lag from G_1 onset to G_2 onset remains constant. In the latter case (represented in Figure 1f), the

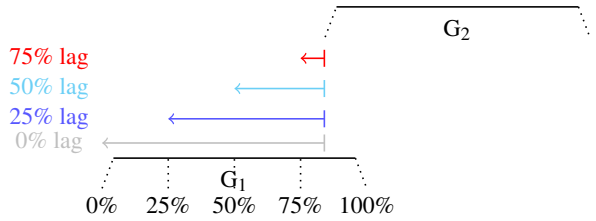

Figure 2: Measuring lag between G_2 and G_1 from different points along G_1

observed lag from G_1 onset to G_2 onset increases, but that from G_1 *offset* to G_2 onset remains constant. That is, in both cases, the observed lag between G_2 onset and a point along the duration of G_1 has the *least variance* when the observed lag is measured from the point along G_1 that G_2 is aligned to.

So, to identify the proportional lag between G_1 and G_2, we can calculate the observed lag between G_2 onset and different proportional points along G_1's duration. This is shown schematically in Figure 2. We can then calculate the variances of the measurements for each of these measured lags to identify the proportional point along G_1 with the lowest variance for the lag measure to G_2 onset.

To test this technique, we created 6 different simulated datasets with random Gaussian noise added to the gesture durations, which differed in the proportional and absolute lags between G_1 and G_2. Specifically, we simulated 3 datasets with an absolute lag of 0 ms that varied in the proportional lag (0% vs. 60% vs. 100% of G_1 duration), and 3 more datasets with an absolute lag of 50 ms and the same three proportional lags. Our approach accurately identified the proportional lags between the two gestures (see Figure 3).

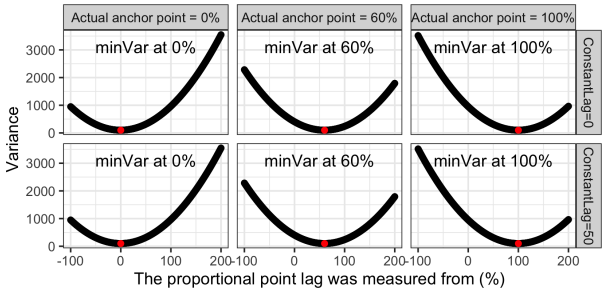

Figure 3: Variances corresponding to different proportional lags on simulated data (red point = proportional lag point with least variance.)

Note, our technique doesn't suffer from the problem of part-whole correlation that plagues the calculation of proportional lag using a ratio of observed lag and G_1 duration [22]. In fact, it is guaranteed to identify the proportional timing relationship between the two gestures as the basic premise of the technique is the statistical theorem: *Variance(X + constant) = Variance(X)* [23]. Essentially, calculating the variances of the observed lags between the two gestures from different points along G_1 allows us to remove the effect of the absolute lag. Note, based on another statistical theorem, it is straightforward to show that the technique works even if the absolute lag has its own variance, as long as the absolute lag and the proportional lag are independent of each other.

3. EXPERIMENTAL DATA

3.1. Methods

In this section, we use the above technique to identify the actual proportional lag in articulatory data. Data were collected from the Wisconsin X-Ray Microbeam Speech Production Database [24], which consists of production data from 57 speakers in three different tasks: wordlists, sentences, and paragraphs. We focussed on words where the initial consonant was a labial, and had multiple repetitions for each speaker with the same preceding context. We further looked for a variety of vowels following the initial consonant. Five words satisfied our requirements: <back, fiber, make, much, people>, where the relevant vowels were [æ, ɑɪ, eɪ, ʌ, i], respectively. Two other words <five, before> came close but were eliminated due to inconsistent contexts. We collected measurements from a total of 789 word productions (back=143,fiber=127, make=171,much=141,people=207).

Lip aperture, defined as the euclidean distance between sensors on the upper and lower lips, was used to track the bilabial gesture; a lower lip ('LL') sensor was used to track the labio-dental gesture; and a tongue blade (either 'T2' or 'T1') sensor was used to track the vocalic gesture. The gestures from each token were parsed using the `findgest()` algorithm in **mview**, a Matlab-based program for data visualization and analysis [25].

For each gesture, we identified the following: (a) gesture onset/offset, identified using a threshold of 10% of the peak maximum, and used, (b) target onset/offset, identified using a threshold of 30% of the peak maximum.

After identifying the onsets and offsets of the gestures/targets, we first calculated the lag duration from 10% steps (range: -50% – 150%) of G_1 duration to the onset of G_2 for the gestural and target measurements, separately. Secondly, for the gestural and target measurements, we calculated the variances for each combination of speaker, word, and step. Finally, we identified the step with the

lowest variance for each combination of speaker and word, also separately for the gestural and target measurements.

3.2. Results

We fitted simple regression models for each of the words, where the dependent variable was the step with the minimum lag for a speaker. Since our interest is simply in identifying the mean step value, each model had only an intercept (therefore, this is equivalent to a one-sample t-test). We also calculated 95% confidence intervals for each of the estimated intercepts. These results are presented in Figure 4. Note, in the interest of space, we don't present tabled summaries, as the crucial information needed for inference is present in the figure.

We focus our discussion on the interpretation of the 95% confidence intervals (CI) for each model [26, 27, 28], instead of just null hypothesis testing of an underlying proportional lag of 0% — this is because our interest here is not only in seeing a consistency (or lack thereof) with the null hypothesis of a proportional lag of 0%, but actually on the range of proportional lag values that are reasonable estimates for the underlying proportional lag. Note, if the CIs don't include a certain proposed lag, that is evidence that the underlying true proportional lag is inconsistent with the proposed lag value.

As can be seen in Figure 4 (left), the onset of the G_2 gesture is not consistent across words/vowels in our data. The 95% CIs suggest the gestural onset of [i] in <people> is consistent with 0% lag (an in-phase alignment with G_1) or with 12.5%-16.7% proportional lag between the onsets of the two gestures (as per the split-gesture hypothesis). However, the 95% CIs of the other four words/vowels are inconsistent with the two hypotheses. Furthermore, two of the three 95% CIs [æ, eɪ] are consistent with the *offset-onset alignment* hypothesis.

In contrast, the 95% CIs in Figure 4 (right) suggest that the proportional timing of onset of the G_2 *target* is consistent only with the offset-onset alignment hypothesis in all five cases.

In fact, when pooled together in a mixed-effects model [29] with a random intercept for speaker and word/vowel, the overall estimated proportional lag for gestural targets was 94% (95% CI: 80% - 107%), which is almost exactly what one would expect as per the offset-onset alignment hypothesis. However, given the by-word/vowel variability for the proportional lags of gestural onsets, it is not surprising that the overall estimated proportional lag for the gestural onsets was 63% with a much wider 95% CI range (33% - 93%).

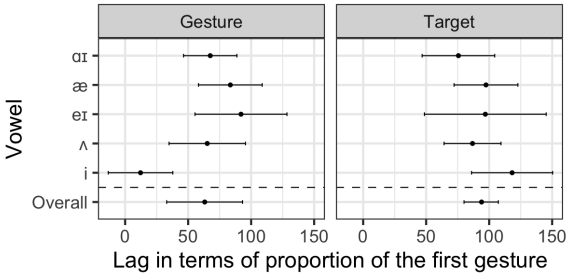

Figure 4: Gestural lag measurements. (points = proportional lag estimates; error bars = 95% CIs)

4. CONCLUSION

In this article, we presented a new technique based on the variance of the observed lag to establish the proportional inter-gestural timing relationship between any two proximate gestures. Using the technique, we observed that vowel gestural onsets were not consistently aligned with the preceding consonant gesture, but the vowel target onsets consistently had offset-onset alignment.

It is important to point out that this is a first attempt using a new technique to establish proportional lag. While the technique is accurate in identifying proportional lags, we caution the reader from over-interpreting our experimental results—while the number of cases here are (at minimum) comparable to previous works looking at the issue either impressionistically or through a careful quantitative study, it is still quite small and is therefore in need of replication.

If replicable and generalisable to other cases, our results have two important implications. First, speakers might actually be planning to coordinate gestural *targets* in a consistent way and not gestural *onsets*. The differences in gestural onsets could be a result of trying to ensure the consistent alignment of gestural targets in a sequential manner. Second, standardly, the C-centre effect noted earlier is seen as a result of competitive coupling wherein consonant gestures in a complex onset are all in an in-phase relationship with the following vowel gesture while the consonant gestures themselves are in an anti-phase relationship. But, if CV timing is an offset-onset relationship then we can't derive the C-centre effect in the standard way. It is, however, possible to derive it from a situation wherein all gestural pairs (CC and CV) in complex onsets have offset-onset timing.

5. REFERENCES

[1] C. P. Browman and L. H. Goldstein, "Articulatory gestures as phonological units," *Phonology*, vol. 6, no. 2, pp. 201–251, 1989.

[2] ——, *Tiers in articulatory phonology, with some implications for casual speech*, ser. Papers in Laboratory Phonology. Cambridge University Press, 1990, vol. 1, pp. 341–376.

[3] L. Goldstein, "Back to the past tense in english," *Representing language: Essays in honor of Judith Aissen*, pp. 69–88, 2011.

[4] Z. Liu, Y. Xu, and F.-f. Hsieh, "Coarticulation as synchronised cv co-onset–parallel evidence from articulation and acoustics," *Journal of Phonetics*, vol. 90, p. 101116, 2022.

[5] C. P. Browman and L. H. Goldstein, "Some notes on syllable structure in articulatory phonology," *Phonetica*, vol. 45, pp. 140–155, 1988.

[6] J. A. Shaw, A. I. Gafos, P. Hoole, and C. Zeroual, "Syllabification in moroccan arabic: evidence from patterns of temporal stability in articulation," *Phonology*, vol. 26, no. 1, pp. 187–215, 2009.

[7] A. Hermes, D. Mücke, and M. Grice, "Gestural coordination of Italian word-initial clusters: The case of 'impure s'," *Phonology*, vol. 30, no. 1, pp. 1–25, 2013.

[8] D. Mücke, A. Hermes, and S. Tilsen, "Incongruencies between phonological theory and phonetic measurement," *Phonology*, vol. 37, no. 1, pp. 133–170, 2020.

[9] H. Nam, "Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure," *Laboratory Phonology*, vol. 9, pp. 483–506, 2007.

[10] E. L. Saltzman and K. G. Munhall, "A dynamical approach to gestural patterning in speech production," *Ecological Psychology*, vol. 1, no. 4, pp. 333–382, 1989. [Online]. Available: https://doi.org/10.1207/s15326969eco0104_2

[11] M. E. Beckman and J. Edwards, "Intonational categories and the articulatory," *Speech perception, production and linguistic structure*, p. 359, 1992.

[12] J. Edwards, M. E. Beckman, and J. Fletcher, "The articulatory kinematics of final lengthening," *The Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 369–382, 1991. [Online]. Available: https://doi.org/10.1121/1.400674

[13] J. A. Shaw, K. Durvasula, and A. Kochetov, "The temporal basis of complex segments," in *Proceedings of the 19th International Congress of Phonetic, Sciences (ICPhS 2019)*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. Australasian Speech Science and Technology Association Inc., 2019, pp. 676–680. [Online]. Available: https://assta.org/proceedings/ICPhS2019/papers/ICPhS_725.pdf

[14] J. A. Shaw, S. Oh, K. Durvasula, and A. Kochetov, "Articulatory coordination distinguishes complex segments from segment sequences," *Phonology*, vol. 38, no. 3, pp. 437–477, 2021.

[15] J. Kelso, E. Saltzman, and B. Tuller, "The dynamical perspective on speech production: data and theory," *Journal of Phonetics*, vol. 14, no. 1, pp. 29 – 59, 1986. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0095447019306084

[16] Z. Liu, Y. Xu, and F. Hsieh, "Coarticulation as synchronised sequential target approximation: An ema study," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2020. International Speech Communication Association (ISCA), 2020, pp. 1381–1385.

[17] A.-P. Benguerel and H. A. Cowan, "Coarticulation of upper lip protrusion in french," *Phonetica*, vol. 30, no. 1, pp. 41–55, 1974.

[18] S. E. Boyce, "Coarticulatory organization for lip rounding in Turkish and English," *The Journal of the Acoustical Society of America*, vol. 88, no. 6, pp. 2584–2595, 1990.

[19] S. E. Boyce, R. A. Krakow, and F. Bell-Berti, "Phonological underspecification and speech motor organisation," *Phonology*, vol. 8, no. 2, pp. 219–236, 1991. [Online]. Available: http://www.jstor.org/stable/4420035

[20] S. Hawkins and S. B. Jones, "The perceptual magnet effect reflects phonetic context," *Journal of the Acoustical Society of America*, vol. 115, no. 5, 2630, 2004.

[21] M.-J. Solé, "Phonetic and phonological processes: The case of nasalization," *Language and Speech*, vol. 35, no. 1-2, pp. 29–43, 1992. [Online]. Available: https://doi.org/10.1177/002383099203500204

[22] W. J. Barry, "Some problems of interarticulator phasing as an index of temporal regularity in speech," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 9, no. 5, 1983.

[23] L. Wasserman, *All of statistics: a concise course in statistical inference*. Springer, 2004, vol. 26.

[24] J. R. Westbury, "X-ray microbeam speech production database user's handbook," 1994.

[25] M. Tiede, "Mview: software for visualization and analysis of concurrently recorded movement data," 2005.

[26] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 236, no. 767, pp. 333–380, 1937.

[27] J. Cohen, "The earth is round (p<. 05)," *American Psychologist,*, vol. 49, no. 12, pp. 997–1003, 1994.

[28] G. Cumming, *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, 2013.

[29] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using `lme4`," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.